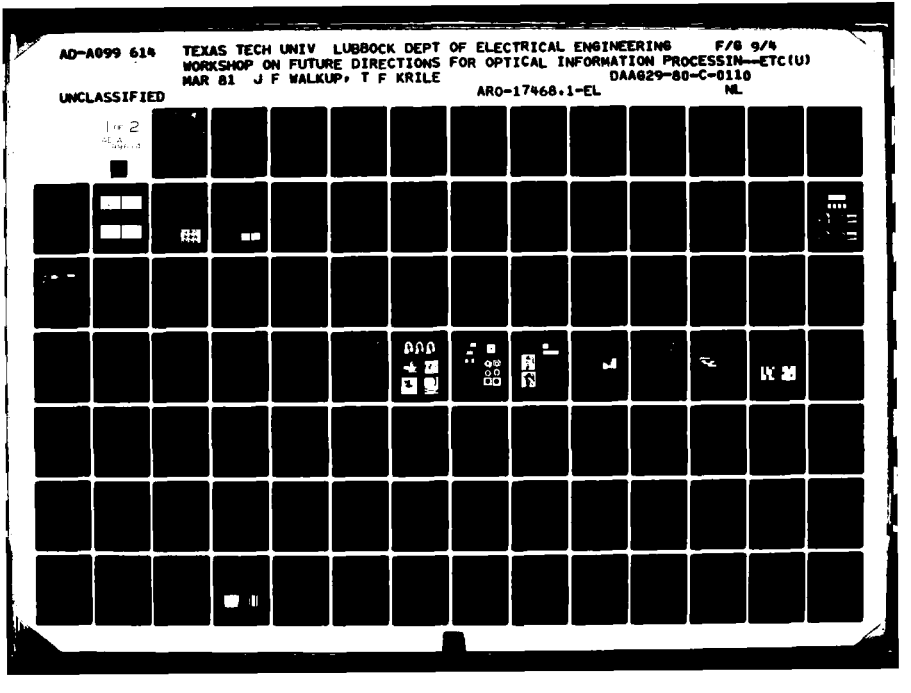
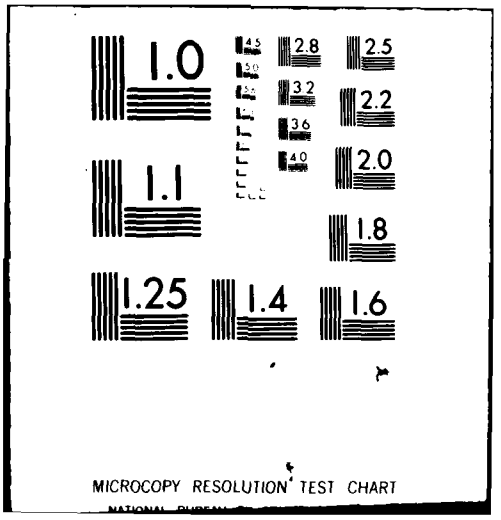

AD-A099 614
TEXAS TECH UNIV LUBBOCK DEPT OF ELECTRICAL ENGINEERING F/G 9/4
WORKSHOP ON FUTURE DIRECTIONS FOR OPTICAL INFORMATION PROCESSIN--ETC(U)
MAR 81 J F WALKUP, T F KRILE DAAG29-80-C-0110
UNCLASSIFIED ARO-17468.1-EL NL
1 of 2


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6
MICROCOPY RESOLUTION TEST CHART

LEVEL II

(18) ARO 17468.1-EL

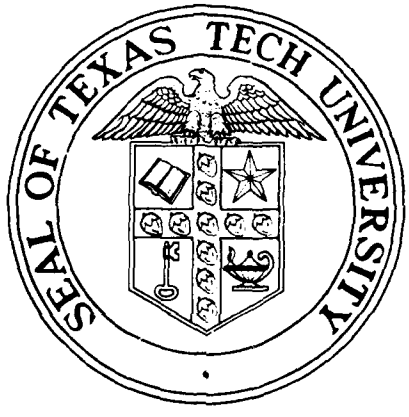
(12) 4

(6) WORKSHOP ON

FUTURE DIRECTIONS FOR OPTICAL INFORMATION PROCESSING

(9) Final Report. 15 Mar 80 - 14 Mar 81

(10) John F. Walkup
Thomas F. Krile



(11) 1 Mar 81

(12) 155

U. S. ARMY RESEARCH OFFICE
CONTRACT DAAG 29-80-C-0110

(15)

DEPARTMENT OF ELECTRICAL ENGINEERING

TEXAS TECH UNIVERSITY

Lubbock, Texas 79409

406840

81 6 01 012

THE FINDINGS IN THIS REPORT ARE NOT TO BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, UNLESS SO DESIGNATED BY OTHER AUTHORIZED DOCUMENTS.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A099614	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) WORKSHOP ON FUTURE DIRECTIONS FOR OPTICAL INFORMATION PROCESSING		5. TYPE OF REPORT & PERIOD COVERED Final: 3/15/80-3/14/81
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) John F. Walkup and Thomas F. Krile		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0110
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Electrical Engineering Texas Tech University Lubbock, Texas 79409		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office P. O. Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE March 1, 1981
		13. NUMBER OF PAGES 152
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE NA

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

optical processing
workshop
information processing
signal processing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report presents the proceedings of the ARO-sponsored workshop on "Future Directions for Optical Information Processing" held at Texas Tech University, Lubbock, Texas, on May 20-22, 1980. Ten papers and a panel discussion consider the state-of-the-art and future areas for research in the field of optical information processing.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Table of Contents

for the Workshop on

FUTURE DIRECTIONS FOR OPTICAL INFORMATION PROCESSING

Department of Electrical Engineering
Texas Tech University
Lubbock, Texas 79409

May 20-22, 1980

*Speaker

Preface

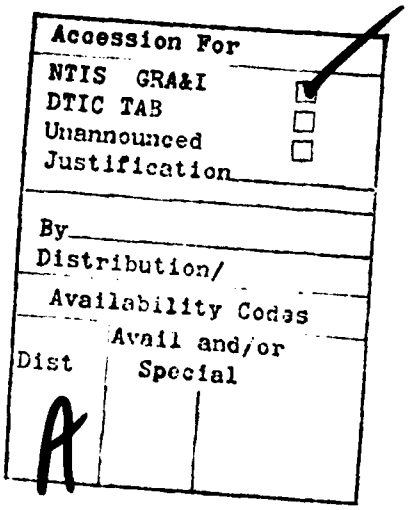
A workshop on the topic "Future Directions for Optical Information Processing" was held on May 20-22, 1980 on the campus of Texas Tech University in Lubbock, Texas. The workshop was sponsored by the U.S. Army Research Office and hosted by Texas Tech's Department of Electrical Engineering. The goals of the workshop were (1) to present a reasonably comprehensive overview of some critical research areas in optical information processing and (2) to discuss which potential areas for future research appeared most attractive.

With these goals in mind, the co-directors put together a program of ten invited speakers and ten discussion leaders. Each speaker also provided a manuscript for these proceedings. The manuscripts are each followed by an edited transcript of the discussion which followed the respective talk. Also presented is an edited transcript of the summary panel discussion which concluded the workshop, and a list of the 37 participants.

A high degree of enthusiasm characterized the presentations, discussions, and the panel discussion. It is hoped that the workshop will be repeated, possibly at two year intervals, with the next workshop tentatively scheduled for 1983.

The co-directors want to acknowledge the financial support provided by the U.S. Army Research Office, with Drs. William A. Sander and Bob D. Guenther acting as contract monitors. The support and assistance of Dean John R. Bradford of the Texas Tech University College of Engineering, Dr. Russell H. Seacat, Chairman of the Department of Electrical Engineering, and the staff and students who assisted with the workshop arrangements are also gratefully acknowledged.

John F. Walkup
Thomas F. Krile
Lubbock, Texas



Space-Variant Coherent Optical Processing

John F. Walkup

Department of Electrical Engineering
Texas Tech University
Lubbock, Texas 79409

Abstract

This paper reviews the state-of-the-art in techniques for performing either one-dimensional (1-D) or two-dimensional (2-D) space-variant operations using coherent light. Examples are presented to illustrate the major categories of processors. Possible future directions which appear promising are suggested and discussed.

1. Introduction

Most of the original interest in coherent optical information processing centered on the possibility of doing linear, space-invariant operations (e.g. convolutions, correlations) using the Fourier transforming properties of converging lenses [1]. In recent years, however, more attention has been devoted to increasing the flexibility of coherent optical processors[2-7]. In particular, considerable attention has been devoted to techniques for performing either one-dimensional (1-D) or two-dimensional (2-D) space-variant operations described by the 1-D and 2-D superposition integrals.

To illustrate, in the 1-D case the superposition integral is given by

$$g(x) = \int_{-\infty}^{\infty} f(\xi)h(x;\xi)d\xi \tag{1}$$

where $f(\xi)$ represents the 1-D input "signal," $h(x;\xi)$ represents 2-D system line spread function or impulse response, and $g(x)$ is the output. When the response $h(x;\xi)$ depends only on $x-\xi$ Eq.(1) becomes the familiar convolution integral,

$$g(x) = \int_{-\infty}^{\infty} f(\xi)h(x-\xi)d\xi \ . \tag{2}$$

In the 2-D input case, the more general linear space-variant operation is described by the superposition integral

$$g(x,y) = \iint_{-\infty}^{\infty} f(\xi,\eta)h(x,y;\xi,\eta)d\xi\ d\eta \tag{3}$$

where $f(\xi,\eta)$ is the 2-D input signal, and $h(x,y;\xi,\eta)$ is now (potentially) a 4-D point spread function describing the system response at output coordinates (x,y) to a point source (impulse) located at input coordinates (ξ,η). Again, in the special case of a space-invariant system, the point spread function becomes $h(x-\xi,\ y-\eta)$ and one obtains the 2-D convolution integral

$$g(x,y) = \iint_{-\infty}^{\infty} f(\xi,\eta)h(x-\xi,\ y-\eta)\ d\xi\ d\eta \tag{4}$$

characteristic of a space-invariant system.

A number of recent review papers and book chapters [2-7] have discussed developments in coherent space-variant optical processing and the applications of these techniques in problem areas such as image restoration, performing various integral transforms in pattern recognition applications, various frequency-variant operations [8], and others. One of the prime motivations for studying techniques for optically evaluating Eqs. (1) and (3) is that in contrast to many temporal signal analysis problems where the systems of interest may be time-invariant, a broad variety of optical systems and optical processing operations are space-variant, including the familiar Fourier transforming and non-unity magnification systems. It should also be noted that while only coherent processors will be discussed here, incoherent processors capable of performing space-variant matrix-vector multiplication operations on either serial [9] or parallel [10] data are also being actively investigated.

2. Techniques for Space-Variant Processing

We will group the techniques to be discussed into four categories: (1) techniques for 1-D inputs; (2) holographic multiplexing techniques; (3) coordinate transformation processing techniques; and (4) other novel techniques.

2.1 Techniques for 1-D Inputs

It has been shown [11,12], based on earlier work by Cutrona [13] that the processor of Figure 1 will perform a generalized 1-D space-variant operation of the form of Eq. (1). As shown the 1-D input $f(\xi)$ is coherently illuminated and placed adjacent to the 2-D processing mask representing the potentially complex kernel $h(x,\xi)$. An astigmatic operation [14] consisting of Fourier transforming along the ξ axis and imaging along the x axis is performed between planes P_1 and P_2 by the 3 cylindrical lenses L_1, L_2, and L_3 (focal lengths related by $2f_1 = f_2 = 2f_3$). In the output plane P_2, the field amplitude $g_o(x,\nu)$ is given by

$$g_o(x,\nu) = \int_{-\infty}^{\infty} f(\xi)h(x;\xi)\exp(-j2\pi\nu\xi)d\xi \tag{5}$$

with $\nu = x/\lambda f_2$ being the scaled spatial frequency. Looking along the $\nu=0$ axis we find that, comparing with Eq. (1),

$$g_o(x,\ 0) = g(x), \tag{6}$$

which indicates that the processor of Fig. 1 can, in principle, perform a broad range of 1-D space-variant operations. Examples of the types of operations which can be performed with the 1-D space-variant processor are numerous. They include variable spatial magnification [11,12], geometrical distortions (coordinate transformations) [4,11,12], and Mellin transforms[3,15-19]. Variations of this processor have also been used to perform operations such as frequency-variant spectral analysis[8]. In addition, the astigmatic lens operation may actually be combined with the $h(x;\xi)$ function to produce a "single optical element" processor consisting of a 1-D input, a processor mask, and a region of free space before the output plane, which indicates some of the power in the processor. The introduction of 1-D time signals using acousto-optical devices offers opportunities for real-time processing subject to device-imposed constraints.

Before leaving the processor of Figure 1 it is well to note that if one makes use of the entire output plane rather than only the $\nu=0$ axis, the processor may be used to evaluate such 2-D functions of a 1-D input as the Laplace transform [21] and the familiar Woodward radar ambiguity function [22-25]. Though these functions are not, strictly speaking, space-variant 1-D operations based on the definition of Eq. (1),[26] they bear mentioning due to the fact that they represent operations of significant interest to the optical information processing community.

Potentially attractive future directions in the area of 1-D space-variant processing include (1) the continued investigation of real time operations using acousto-optical devices; (2) exploring further the use of computer-generated holographic masks (CGH's) to represent $h(x;\xi)$ for a variety of integral transforms and other processing operations; and (3) investigating the use of 1-D processors plus time-integrating detectors to perform 2-D space-variant operations of the form of Eq. (3). It is clear that the class of 1-D space-variant operations which should be performable is almost certainly much larger than has been explored to date.

2.2 Holographic Multiplexing Techniques

When we move from 1-D space-variant processors to the 2-D processors described in Eq. (3), new approaches must be considered. The problem is that $h(x,y;\xi,\eta)$ may, in principle, be a different function of (x,y) for each input point (ξ,η). One potentially attractive class of approaches is to multiplex holograms describing the responses of the space-variant system to various inputs. At Texas Tech University we have been considering three such multiplexed hologram approaches: (1) a sampling theorem-based approach; (2) a piecewise isoplanatic approximation (PIA)-based approach; and (3) a discrete orthonormal basis set approach.

The sampling theorem-based approach makes use of a modification, for space-variant systems, of the familiar Whittaker-Shannon sampling theorem [27,28]. We assume that $h(x,y;\xi,\eta)$ varies in a bandlimited manner as a function of (x,y) for variations in the input coordinates (ξ,η). The modified sampling theorem states that the system's output can be perfectly reconstructed (in the absence of noise) by sampling the input plane at a rate proportional to the sum of the input's bandwidth and a "variation bandwidth" which characterizes how rapidly $h(x,y;\xi,\eta)$ changes with ξ and η. For space-invariant systems this variation bandwidth goes to zero as expected.

Another multiplexed hologram approach, the PIA approach [29-31] would store only one hologram per isoplanatic patch, and would thus, in principle, be input bandwidth-independent as far as the number of stored holograms required. The third approach mentioned - i.e. the discrete orthonormal basis set approach - would store the system's responses to each element of an orthonormal basis set spanning the space of possible inputs [29,31]. Here the number of stored holograms would be system-independent but would be input-dependent. To date the approach investigated in most detail has been the sampling theorem-based approach.

Figure 2 illustrates the basic idea behind the sampling theorem-based approach. The input plane is coherently illuminated and spatially sampled. After a Fourier transforming operation by lens L_2, the many plane waves associated with different input plane samples access holographic filters multiplexed in the single holographic medium shown placed in the Fourier plane. We assume that each filter is accessed so that

crosstalk with other filters is suppressed. Then, assuming coherent addition of waves in the output plane after another Fourier transforming operation by lens L_3, the output will closely approximate the desired output-i.e. a discrete approximation to the superposition integral of Eq. (3). The approach of Fig. 2 is attractive since in principle, the space-bandwidth product of the processor is limited only by our ability to holographically multiplex the filters.

An early approach to suppressing hologram-to-hologram crosstalk in the sampling theorem-based approach was to employ the Bragg extinction angle condition associated with thick recording media [32]. Unfortunately the Bragg phenomenon is not nearly as effective in suppressing crosstalk when sampling 2-D inputs as it is in sampling 1-D inputs.

A more attractive approach to the multiplexed hologram processor involves the use of phase-coded reference beams [33-38]. An advantage of this approach is that thin recording media may be used, although additional crosstalk suppression may be had using thick media. The setups for recording and playing back a multiplexed hologram processor for multiple phase-coded reference beams are indicated in Figs. 3 and 4. As shown in Fig. 3, one records the multiplexed processor by sampling the space-variant system's input plane with the i^{th} sample producing point spread function h_i. The i^{th} reference point source simultaneously illuminates the i^{th} member of a family of n phase-encoding diffusers (e.g. shower glass, ground glass, binary-valued Gold code masks, etc.). Lenses L_1 and L_2 perform Fourier transform operations, producing sequentially recorded holograms of the m respective transfer functions of the space-variant system, with each reference beam uniquely phase-encoded. When one simultaneously plays back all n holograms, as in Fig. 4, by spatially sampling an input s, the output 0 is mathematically described by

$$0 = \sum_{i=1}^{n} s_i \, h_i * (\tilde{M}_i \bigstar \tilde{M}_i) + \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} s_i \, h_j * (\tilde{M}_i \bigstar \tilde{M}_j) \tag{7}$$

where * represents convolution and $\bigstar$ represents correlation. Since the ideal output is given by

$$0 = \sum_{i=1}^{n} s_i \, h_i \tag{8}$$

it is clear that we seek a family of "effective"diffuser functions (the products of the transmittances of the diffusers and the complex amplitudes of the illuminating waves) such that all the autocorrelations $(\tilde{M}_i \bigstar \tilde{M}_i)$ are essentially Dirac delta functions, and all the crosscorrelations $(\tilde{M}_i \bigstar \tilde{M}_j)$ are essentially zero.

To illustrate the discussion just presented, Figure 5 shows the output of a multiplexed holographic processor representing a 1.5x magnifier for a 2x2 input array, with each diffuser (ground glass) section illuminated with a plane wave [35, 37]. The n(n-1) = 4(3) = 12 crosstalk terms have been distributed into the noise-"balls" surrounding each of the desired n=4 output points. Figure 6 illustrates the fact that "chirping" the diffusers by illuminating them with spherical wavefronts rather than plane wavefronts produce "effective" diffusers which are superior to those produced by plane wave illumination [34-37].

Figure 7 shows the optical output of a badly distorted magnifier designed to blur each of 13 input points in a different manner (clearly space-variant). Figure 8 then shows the output of the holographic processor designed to replace the system which produced the output of Fig. 7. Here we accessed 13 holograms in a stored array (5x5) of 25 holograms, with "chirped" illumination being used. The diffuse crosstalk is clearly visible, as are the effects of vignetting (fact that some of the "tips" of the "H" are missing).

Recent analytical work [36] has shown that binary (2-level) phase diffusers can, in principle, perform as well as multi-level phase diffusers as long as all allowable phase levels are equiprobable. Other work has shown the value of computer-multiplexed holograms in permitting one to avoid the bias buildup problems associated with the direct recording of n multiplexed holograms [38]. Preliminary results, using multiplexed, coded computer generated holograms for space-variant processing, and in matched filtering for pattern recognition [39], have been encouraging, though more work is needed to fully explore this class of techniques.

An alternative approach to multiplexing holograms for the purpose of space-variant processing is based on sampling both the system's input plane and the Fourier plane, as shown in Fig. 9 [40]. Here we assume the system point spread functions to be approximately space-limited in addition to the earlier assumption that the input plane is spatially sampled. This additional assumption permits one to sequentially spatially sample the respective transfer functions using a movable binary mask with periodic openings, as shown in Fig. 9. The playback setup is indicated in Fig. 10. The clear advantages of this technique over the technique described earlier in Figs. 3 and 4 are that (1) only one reference beam needs to be used and (2) there is, in principle, zero crosstalk on playback due to the fact that the Fourier plane samples are spatially nonoverlapping. The disadvantage lies in the fact that, as shown in Fig. 10, the input function must be sampled and replicated spatially (i.e. a multiple input function transparency, etc.) on playback.

Figures 11 and 12 illustrate the results of an experiment designed to test the concept [40]. Nonoverlapping impulse responses h_1, ..., h_4 are shown along with their sum. A composite computer-generated hologram (CGH) corresponding to the sampled transfer functions corresponding to h_1, ..., h_4 was generated and played back, with the experimental result shown in Fig. 12. It was also shown that individual point spread functions could be accessed, and that the technique is not limited to nonoverlapping point spread functions. The fact that coherent addition is taking place was verified. A method for reducing quantization errors in

generating computer-multiplexed holograms by premultiplying the point-spread function by a binary "checkerboard" phase function was also investigated in conjunction with this technique.

2.3 Coordinate Transformation Processing Techniques

This is one area where earlier developments in space-variant digital image processing [41,42] stimulated related work in coherent optical processing. Several approaches have been investigated. One class is described in block diagram form in Figure 13. Assuming a 2-D input $f(\xi,\eta)$, the first step is to perform an invertible coordinate transformation to yield the new function $\hat{f}(\hat{\xi},\hat{\eta})$. Next one performs a linear space-invariant filtering operation on $\hat{f}(\hat{\xi},\hat{\eta})$ to yield the function $\hat{g}(\hat{x},\hat{y})$. The second coordinate transformation is the inverse of the first, and produces the processed output $g(x,y)$. The Abel transform [15], which arises when circularly symmetrical distributions in two dimensions are projected in one dimension is one example of a space-variant processing operation which may be performed as indicated in Fig. 13.

A second type of space-variant coordinate transformation processing technique is shown in Figure 14. Here the input $f(\xi,\eta)$ is coordinate-transformed to yield $\hat{f}(\hat{\xi},\hat{\eta})$. This initial coordinate transformation is chosen to enable one to produce the desired output $g(x,y)$ using a simpler, or more easily performable space-variant filtering operation. A well-researched example is the complex Mellin transform [3,6,19] where, by taking the natural logarithms of the input coordinates, one is able to then obtain the Mellin transform by taking a simple Fourier transform of the coordinate-transformed input function. Though the Fourier transform is still a space-variant operation, it is a straightforward one to obtain. Casesent and his colleagues have used the Mellin transform to perform scale-invariant coherent pattern recognition, along with a number of other applications [16,17,43-48].

A final approach worth mentioning here is the use of phase plates or computer-generated holograms (CGH's) to produce a spatially-varying coordinate transformation. This approach is based on a ray-optics theory of the simultaneous control of the position and direction of rays. This technique, due to Bryngdahl [49], is described in detail in several references. The CGH acts as a generalized grating, with the grating's spatial frequency being a function of the input coordinates-hence the space-variant action. The technique is limited by the requirement that the input data must be spatially coarse when compared to the structure of the phase plate or CGH. This limits the space-bandwidth product of optical processors designed using the technique for coordinate transformation processing. The reader is referred to a recent book chapter by W. H. Lee [50] for a review of this and other optical processing applications of CGH's.

2.4 Other Novel Techniques

The techniques described above do not, by any means, exhaust all possibilities for performing coherent space-variant processing. Carlson and Francois [51,52] have shown that an N+1 plane processor consisting of alternating multiplicative operations and 2-D Fourier transforming operations can perform a variety of linear operations, including space-variant operations. This processor is sketched in Figure 15.

The optical setup for a polychromatic space-variant processor is shown in Figure 16 and is being investigated by Strand and Sawchuk [53]. A polychromatic collimated source illuminates an input plane transparency. A color-multiplexed spatial filter is placed in the Fourier plane. Each filter affects a different wavelength region of the illumination. In the output plane one places a second color mask which can select any one or a combination of the spatially filtered outputs at each point of the output plane. This output mask then determines the nature of the system's space-variant characteristic. Alternatively, a color-encoding mask can be placed adjacent to the object in the input plane. Figure 17(b) illustrates the results obtained when the word "SPACE" in 17(a) was coded red and low-pass filtered, whereas the word "VARIANT" was coded blue and high-pass filtered. An advantage of this technique is the fact that it imposes no severe restrictions on the space-bandwidth product of the input.

A novel technique using volume-hologram multiplexed filters has recently been suggested by Case [54] based on earlier work by Friesem and Peri [55]. As shown in Figure 18 an input object O is coherently illuminated, with a volume hologram filter F placed directly behind the input plane. The lens L images the transmitted light onto the image plane I. Since the higher spatial frequency components of the input O are diffracted at larger angles than lower spatial frequency information, and since many holograms could be multiplexed spatially in the medium using the Bragg extinction angle phenomenon discussed earlier, definite possibilities exist for using this input plane Fourier processing scheme to do space-variant processing. Again we expect that the limitations on our space-bandwidth processing capability will be set by the practical limitations associated with storing angle-multiplexed holograms in a thick recording medium, but the technique appears worthy of additional investigation.

The final novel technique we'll mention might be referred to using the term "temporal holography." It makes use of the temporal integrating and summing properties of a hologram [56,57]. As a result, when the hologram is played back, information concerning the time integral of the amplitude and phase variations of an object wave may be recovered. As indicated in Figure 19, a 2-D spatial light modulator has been replaced by a scanning beam [57] which is temporally amplitude-and/or phase-modulated. The linear optical processor (possibly space-variant) processes the scanned input and the recording medium records a hologram of the scanned input and the reference beam shown. On playback, as shown in Figure 20, one of the terms is proportional to the time-integral of the product of the scanning input and the point-spread function of the linear processor. The low pass filtering in the y-direction [57], based on the sampling theorem, acts to produce a

better approximation to the desired system output. It has been shown that this approach permits one to realize a discrete approximation to the superposition integral [56,57].

Figure 21 indicates how one could, in principle, use temporal holography to do space-variant coherent processing based on sampling the input plane of a processor [56]. If, for the nm^{th} input sample (i.e. the sample located at (ξn, ηm) as shown) we introduce the $H_{nm}(f_x,f_y)$ transfer function in the Fourier plane, then we will record a temporal hologram containing the superposition of the product of each input and its respective point spread function. One potential advantage to this temporal-holography approach to space-variant processing is that the processor to be represented need not be physically realizable (i.e. we only need to have the collection of $H_{nm}(f_x,f_y)$ masks desired). Potential disadvantages include the need for sequential data inputs, and the need to change the Fourier plane "masks" for each input sample. We note, however, that the piecewise isoplanatic approximation approach and the discrete orthonormal basis set approach could also be applied here, and these might reduce the number of required "masks."

3. Space-Variant Processing: Future Directions

It is clear that the many applications for space-variant optical processors are sufficient to stimulate research interests in the field for some time to come. Areas needing additional work include (a) the need to increase the space-bandwidth products of the processors; (b) the need to handle inputs and processing in real time; and (c) the need to use the well understood 1-D space-variant processing systems more widely, possibly including their use in 2-D processors. In addition, from some of the recent work in incoherent optical processing, the area of incoherent space-variant processors using multiple parameters, including color and/or polarization to perform operations on real and complex inputs appears to be an attractive one for future work.

4. Acknowledgements

The author wishes to acknowlege the many useful discussions with T. F. Krile, M. O. Hagler, and R. J. Marks II during the preparation of this paper. The assistance of R. Kasturi with the figures and Ms. Heidi Hanssen with the typing of the manuscript is also acknowledged. The portion of the research discussed which was performed at Texas Tech University has been supported by the Air Force Office of Scientific Research under AFOSR grants 75-2855 and 79-0076.

5. References

1. J. W. Goodman, Introduction to Fourier Optics (McGraw-Hill, New York, 1968).

2. J. W. Goodman, "Operations Achievable with Coherent Optical Information Processing Systems," IEEE Proc. 65, 29-38 (1977).

3. J. W. Goodman, "Linear Space-Variant Optical Data Processing," (to appear in Optical Data Processing: Fundamentals, ed. by S. Lee, Springer-Verlag, 1980).

4. D. Casasent and D. Psaltis, "New Optical Transforms for Pattern Recognition," IEEE Proc. 65, 77-84 (1977).

5. D. Casasent, "Optical Signal Processing," (pp. 241-282) in Optical Data Processing: Applications, ed. by D. Casasent, Vol. 23, Springer-Verlag. "Topics in Applied Physics," (Springer-Verlag, Berlin, 1978).

6. J. F. Walkup, "Space-Variant Coherent Optical Processing," Opt. Engr., 19, 339-346 (1980).

7. W. T. Rhodes, "Space-Variant Optical Systems and Processing," in Applications of the Optical Fourier Transform, H. Stark, ed. (Academic Press, New York, 1981).

8. W. T. Rhodes and J. M. Florence, "Frequency Variant Optical Signal Analysis," Appl. Optics 15, 3073-3079 (1976).

9. M. A. Monahan, K. Bromley, and R. P. Bocker, "Incoherent Optical Correlators," IEEE Proc. 65, 121-129 (1977).

10. J. W. Goodman et. al. - see paper in this proceedings.

11. J. W. Goodman, P. Kellman, and E. W. Hansen, "Linear Space-Variant Optical Processing of 1-D Signals," Appl. Optics 16, 733-738 (1977).

12. R. J. Marks II, J. F. Walkup, M. O. Hagler, and T. F. Krile, "Space-Variant Processing of 1-D Signals," Appl. Optics 16, 739-745 (1977).

13. L. J. Cutrona, "Recent Developments in Coherent Optical Technology," in Optical and Electro-Optical Information Processing, J. T. Tippett et. al. Eds. (MIT press, Cambridge, 1965), Chapter 6.

14. R. J. Marks II and S. V. Bell, "Astigmatic Coherent Processor Analysis," Opt. Engr. 17, 167-169 (1978).

15. R. N. Bracewell, The Fourier Transform and its Applications (McGraw-Hill, New York, 1965).

16. D. Casasent and D. Psaltis, "Doppler Signal Processing: A New Technique," Appl. Opt. 15, 2015 (1976).

17. D. Casasent and C. Szczutkowski, "Optical Mellin Transforms Using Computer Generated Holograms," Opt. Comm. 19, 217-222 (1976).

18. P. Kellman and J. W. Goodman, "Coherent Optical Implementation of 1-D Mellin Transforms," Appl. Opt. 16, 2609-2610 (1977).

19. R. J. Marks II and J. N. Larson, "One-Dimensional Mellin Transformation Using a Single Optical Element," Appl. Opt. 18, 754-755 (1979).

20. R. J. Marks II, M. I. Jones, E. L. Kral and J. F. Walkup, "One-Dimensional Linear Coherent Processing Using a Single Optical Element," Appl. Opt. 18, 2783-2786 (1979).

21. M. R. Mueller and F. P. Carlson, "Bandlimiting Effects in an Optical Laplace Transform Computer," Appl. Optics 14, 2207-2212 (1975).

22. R. A. K. Said and D. C. Cooper, "Crosspath Real-Time Optical Correlator and Ambiguity-Function Processor, Proc. IEE (London), 120, 423-428 (1973).

23. R. J. Marks II, J. F. Walkup, and T. F. Krile, "Ambiguity Function Display: An Improved Coherent Processor," Appl. Opt. 16, 746-750 (1977); addendum 16, 1777 (1977).

24. R. J. Marks II and M. W. Hall, "Ambiguity Function Display Using a Single 1-D Input," Appl. Opt. 18, 2539-2540 (1979).

25. D. P. Casasent and B. V. K. Vijaya Kumar, "Optical Image Plane Correlator for Ambiguity Surface Computation," Appl. Opt. 18, 1673-1678 (1979).

26. B. E. A. Saleh, "Bilinear Processing of 1-D Signals By Use of Linear 2-D Coherent Optical Processors," Appl. Opt. 17, 3408-3409 (1978). Saleh points out that the ambiguity function is actually a form of bilinear transformation.

27. R. J. Marks II, J. F. Walkup, and M. O. Hagler, "A Sampling Theorem for Space-Variant Systems," J. Opt. Soc. Am. 66, 918-921 (1976).

28. R. J. Marks II, J. F. Walkup, and M. O. Hagler, "Sampling Theorems for Linear Shift-Variant Systems," IEEE Trans. on Circuits and Systems CAS-25, 228-233 (1978).

29. R. J. Marks II, J. F. Walkup, and M. O. Hagler, "Volume Hologram Representations of Space-Variant Systems," Proc. International Conf. on Applics. of Holography and Opt. Data Processing, Jerusalem, Aug. 1976, Pergamon Press, Oxford, 1977, 105-113.

30. R. J. Marks II and T. F. Krile, "Holographic Representation of Space-Variant Systems: System Theory," Appl. Opt. 15, 2241-2245 (1976).

31. R. J. Marks II, J. F. Walkup, and M. O. Hagler, "Methods of Linear System Characterization Through Response Cataloging," Appl. Opt. 18, 655-659 (1979).

32. L. M. Deen, J. F. Walkup, and M. O. Hagler, "Representations of Space-Variant Optical Systems Using Volume Holograms," Appl. Opt. 14, 2438-2446 (1975).

33. T. F. Krile, R. J. Marks II, J. F. Walkup, and M. O. Hagler, "Holographic Representations of Space-Variant Systems Using Phase-Coded Reference Beams," Appl. Opt. 16, 3131-3135 (1977).

34. T. F. Krile, M. O. Hagler, W. D. Redus, and J. F. Walkup, "Multiplex Holography with Chirp-Modulated Binary Phase-Coded Reference-Beam Masks," Appl. Opt. 18, 52-56 (1979).

35. M. I. Jones, J. F. Walkup, and M. O. Hagler, "Multiplex Holography for Space-Variant Optical Computing," SPIE 177, 16-21 (1979).

36. E. L. Kral, J. F. Walkup, and M. O. Hagler, "Correlation Properties of Diffusers for Multiplex Holography," J. Opt. Soc. Am. 69, 1409 (1979) (abstract).

37. M. I. Jones, "Multiplex Holography for Two-Dimensional Space-Variant Optical Processing," M. S. thesis, Dept. of Electrical Engineering, Texas Tech University, Aug. 1979.

38. C. A. Irby, M. O. Hagler, and T. F. Krile, "Computer-Generated Multiplex Holography," J. Opt. Soc. Am. 69, 1409 (1979) (abstract).

39. S. H. Lee, private communication, Aug. 1979.

40. R. Kasturi, T. F. Krile, and J. F. Walkup, "Multiplex Holography for Space-Variant Processing: A Transfer Function Sampling Approach," Appl. Opt., 20, 881-886 (1981).

41. G. M. Robbins and T. S. Huang, "Inverse Filtering for Linear Shift-Variant Systems," IEEE Proc. 60, 862-872 (1972).

42. A. A. Sawchuk, "Space-Variant Motion Degradation and Restoration," IEEE Proc. 60, 854-861 (1972).

43. D. Casasent and D. Psaltis, "Position, Rotation, and Scale Invariant Optical Correlation," Appl. Opt. 15, 1795-1799 (1976).

44. D. Casasent and D. Psaltis, "Deformation Invariant, Space-Variant Optical Pattern Recognition," in Progress in Optics, Vol. XVI, E. Wolf, Ed., pp. 289-356. (North-Holland, Amsterdam, 1978).

45. D. Casasent and D. Psaltis, "Multiple-Invariant Space-Variant Optical Processors," Appl. Opt. 17, 655-659 (1978).

46. D. Casasent and M. Kraus, "Polar Camera for Space-Variant Pattern Recognition," Appl. Opt. 17, 1559-1561 (1978).

47. D. Casasent and D. Psaltis, "Waveform Considerations in Space-Variant Optical Processors," Opt. Letters 4, 18-20 (1979).

48. D. Casasent and D. Psaltis, "Space-Bandwidth Product and Accuracy of the Optical Mellin Transform," Appl. Opt. 16, 1472 (1977).

49. O. Bryngdahl, "Geometrical Transformations in Optics," J. Opt. Soc. Am. 64, 1092-1099 (1974).

50. W. H. Lee, "Computer-Generated Holograms: Techniques and Applications," in Progress in Optics, Vol. XVI, E. Wolf Ed. pp. 119-232 (North-Holland, Amsterdam, 1978).

51. F. P. Carlson and R. E. Francois Jr., "Generalized Linear Processors for Coherent Optical Computers," IEEE Proc., 65, 10-18 (1977).

52. R. E. Francois, Jr. and F. P. Carlson, "Iterative Fourier Approach for Describing Linear, Multiple Plane, Coherent Optical Processors," Appl. Opt., 18, 2775-2782 (1979).

53. T. C. Strand and A. A. Sawchuk, "Space-Variant Processing with Polychromatic Light," Proc. of ICO-11 Conference, Madrid, Spain, 1978 (pp. 269-272), J. Bescos, et. al. Eds.

54. S. K. Case, "Fourier Processing in the Object Plane," Opt. Letters 4, 286-288 (1979).

55. D. Peri and A. A. Friesem, "Image Restoration Using Volume Diffraction Gratings," Opt. Letters 3, 124-126 (1978).

56. R. J. Marks II, "Two-Dimensional Coherent Space-Variant Processing Using Temporal Holography: Processor Theory," Appl. Opt., 18, 3670-3674 (1979).

57. M. O. Hagler, R. J. Marks II, E. L. Kral, J. F. Walkup, and T. F. Krile, "Scanning Technique for Coherent Processors," Appl. Opt., 19, 4253-4257 (1980).

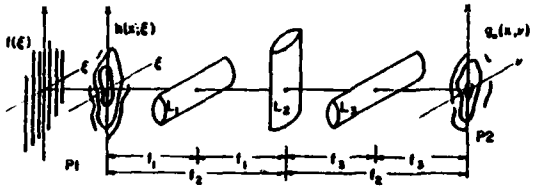


Fig. 1. A coherent optical processor for performing 1-D space-variant operations (Ref. 12).

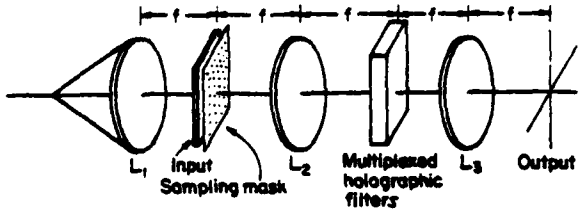


Fig. 2. Basic idea for 2-D space-variant processor using holographically multiplexed filters in the Fourier plane of a coherent processor.

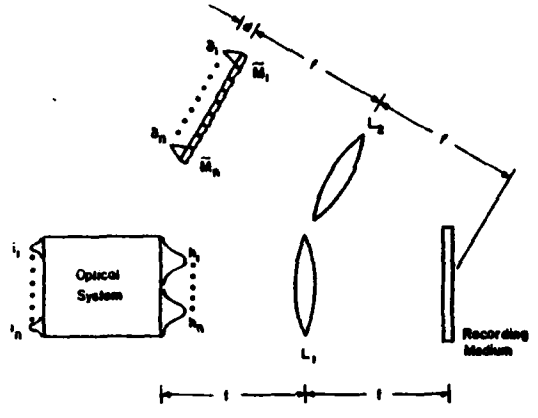


Fig. 3. Setup for sequentially recording n multiplexed holograms using n phase-coded reference beams (Ref. 34).

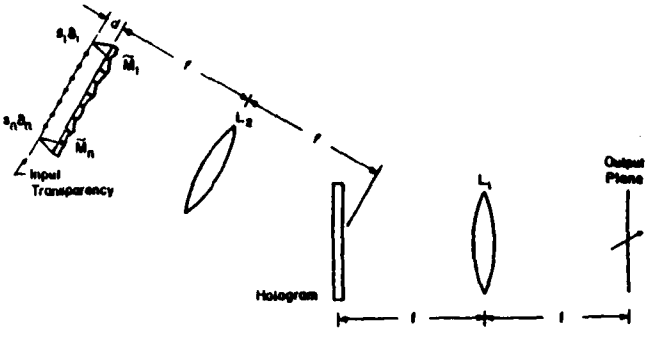


Fig. 4. Setup for playback of the multiplexed holographic processor of Fig. 3 (Ref. 34).

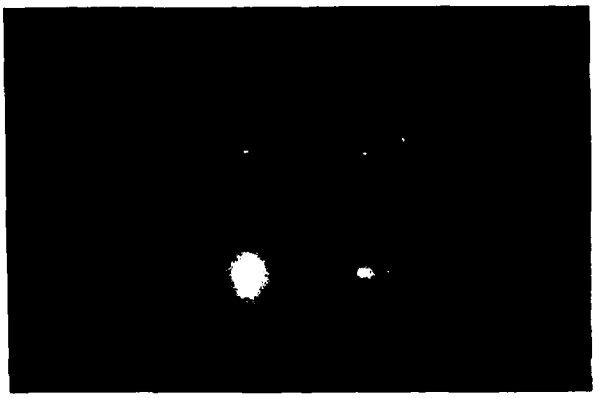


Fig. 5. Output of holographic processor for 2x2 input array, 1.5x magnifier, ground glass diffuser and plane-wavefront diffuser illumination (Refs. 35,37).

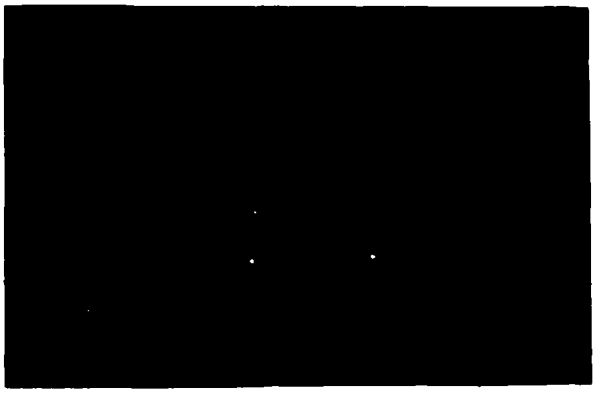


Fig. 6. Output of processor analogous to that of Fig. 5, but using spherical wavefront ("chirped") diffuser illumination (Refs. 35,37).

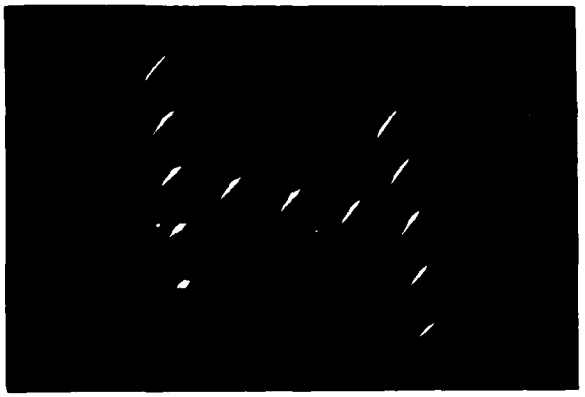


Fig. 7. Optical output of a badly distorted mangifier for the 13 input points shaped like the letter "H" (Refs. 35,37).

Fig. 8. Output of the holographic processor designed to replace the system used to produce Figure 7 (Refs. 35,37).

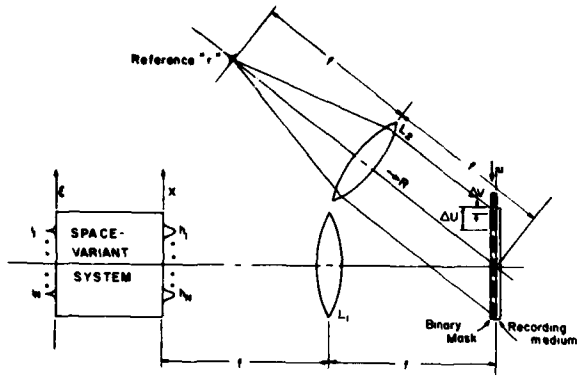


Fig. 9. Optical recording scheme for the sampled input plane/sampled Fourier plane multiplexed holographic processor (Ref. 40).

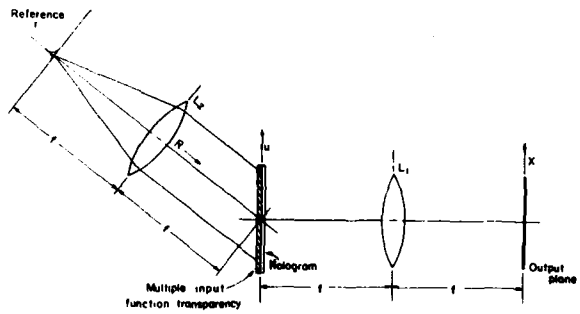


Fig. 10. Scheme for playing back the processor recorded in Figure 9 (Ref. 40).

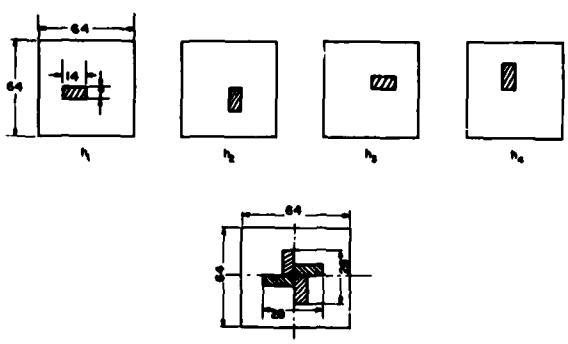


Fig. 11. Four disjoint impulse responses and their sum used in experiment to test the scheme of Figures 9-10 (Ref. 40).

Fig. 12. Experimental optical output, corresponding to Fig. 11, with all four impulse responses accessed (Ref. 40).

Fig. 13. The general form for one coordinate transformation approach to obtaining a space-variant processor.

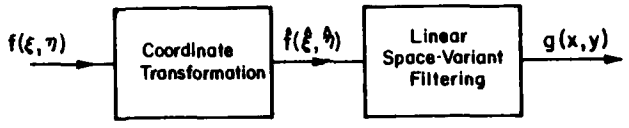


Fig. 14. A second type of coordinate transformation processor.

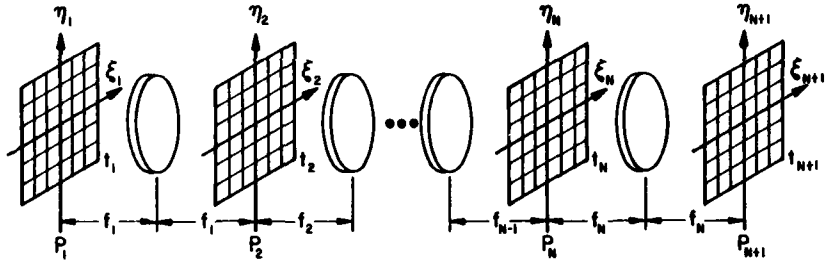


Fig. 15. An N+1 plane generalized linear coherent processor (Refs. 51,52).

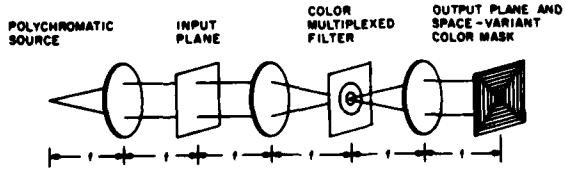


Fig. 16. Scheme for performing space-variant processing using polychromatic light (Ref. 53).

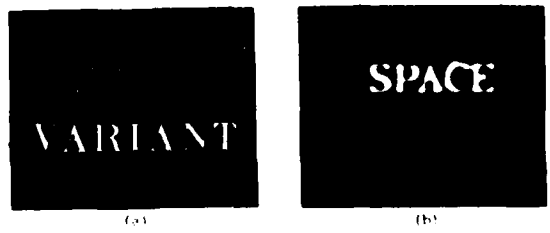


Fig. 17. Space-variant processing using the setup of Figure 16. Here (a) is the input object and (b) is the output. The upper word has been low-pass filtered and the lower word has simultaneously been high-pass filtered (Ref. 53).

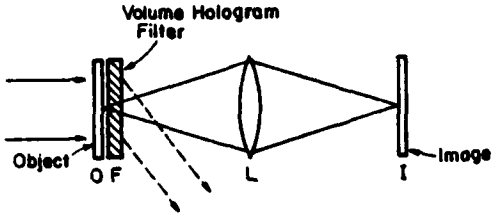


Fig. 18. Technique for space-variant processing using a multiplexed volume hologram processor in the object plane (Ref. 54).

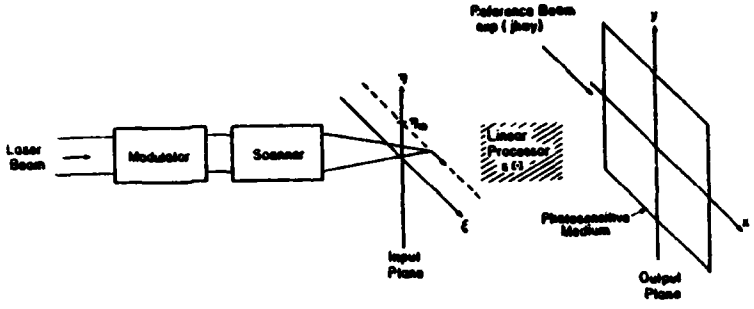


Fig. 19. Input scanning technique for recording a hologram of the temporally modulated scanning beam processed by a linear processor (possibly space-variant). (Ref. 57)

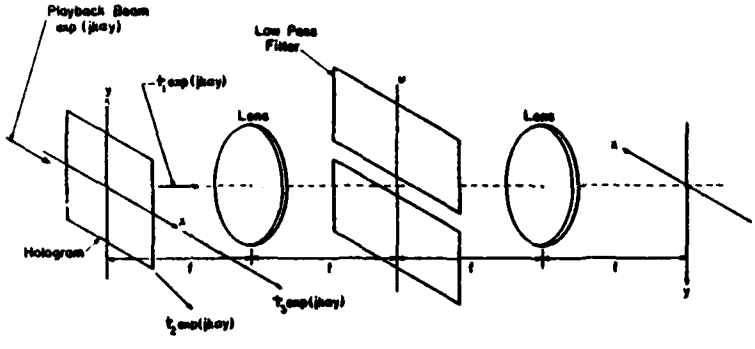


Fig. 20. Scheme for playing back the hologram recorded in Fig. 19 (Ref. 57).

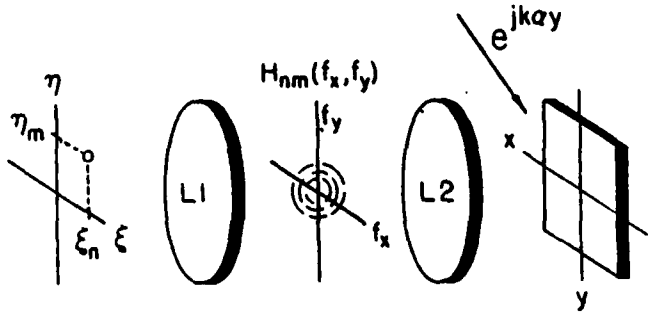


Fig. 21. Sampling theorem-based approach for using temporal holography to perform a 2-D space-variant processing operation (Ref. 56).

Discussion (John F. Walkup; Discussion Leader: Robert J. Marks II)

[Key: Q = Question, A = Answer, C = Comment, and R = Response.]

Note: A review article by Dr. Walkup covering some of this material appears in the May/June 1980 issue of Optical Engineering.

C. There has been a lot of emphasis on space-variant processing looking for general solutions, and it seems to me that frequently where general solutions are necessary, the systems have to adapt to different inputs. A good example is space-variant noise reduction where you have signal dependent noise. In a certain sense you have to know what your signal is before you can design your processor, so general solutions may present a particularly difficult problem. The adaptive nature of optical systems is an area of considerable difficulty at this time. Perhaps the rewards would be as great for looking for solutions to specific situations where the space-variance is well known as in looking for a general processor which will handle those specific situations.

R. That's a good point. There may also be some profit in cataloging the number of specific situations you can get solutions for, and then seeing if there's any way to generalize some of them.

C. Generalized processor schemes tend to water things down, and there's the potential of losing some of the processing power by doing that.

R. In the 1-D case, a general solution exists. The system Dr. Goodman will talk about later can do arbitrary 1-D space-variant relations.

C. That's why I'm intrigued about exploiting the 1-D processor, especially for 2-D operations, because we can do the 1-D well.

Q. Is the 1-D processor going to have the potential to compete with digital processors? If you take a 1-D space-variant operation and digitize it, you get a matrix-vector multiplier, and that seems to be a thing that could be done very quickly digitally.

A. I think Joe Goodman will talk about that, along with Keith Bromley's work. I was really impressed with that because, with discrete inputs, he can do a space-variant operation.

C. A point of nomenclature: a space-variant system that adapts to a particular signal is actually a nonlinear processor, it's no longer a space-variant linear processor.

C. You mentioned that the volume hologram approach has the disadvantage of responding to an entire cone of angles when you'd like it to respond to just one point. Also, the phase-codes multiplexing technique has a buildup of background noise "halo." Perhaps there's fruitful work to be done in combining these two approaches.

R. When Mike Jones was working in our group, he noticed some volume effects helping him out when he went to thicker media in the phase-coded approach, so I think you're right.

C. You may buy something good by using the 1-D angular selectivity of a volume hologram in one dimension and then a 1-D diffuser in the other dimension.

C. When you talk about separable kernels to do 2-D operations with 1-D processors, there are a great many transforms that are of practical interest and are space-variant, such as the Walsh-Hadamard transform.

R. I agree. At first it seemed to us as if the inputs had to be separable as well, and that would be pretty restrictive. But there are a lot of situations of interest with separable kernels.

C. Casasent and Psaltis have an interesting paper in the last issue of Optical Engineering on different architectures for optical processing, including space-variant processing.

C. Just as an aside, the angular selectivity of volume holograms in one direction and the lack of it in the orthogonal direction can be a big advantage where you want to have multiple beams accessing the hologram.

Q. What are some of the more important applications people have in mind for 2-D space-variant processors?

A. Restoration of blurred images where you have a space-variant blur, such as panning cameras, is one motivation.

Q. Is this a continuing problem in, say, aerial imaging?

A. Not in the better systems, but I'm sure there are other applications for blur removal or transform modification.

C. In many vehicles they have unusually shaped windows in order to fit the window to the vehicle, and they

don't have circular symmetry.

Q. Don't you have to talk about real-time processing there?

R. Yes, you have to talk about that all the time.

Q. In these systems which work in the lab environment we're processing images having only a few resolution cells, typically we display on 500 pixel by 500 line monitors. But in the real world the goal is to process 5,000 to 10,000 pixels in each dimension. Will we be able to extend our lab systems to this kind of resolution, or will we be forever limited to rather small images?

A. A lot of these approaches are presently somewhat limited as to space-bandwidth product, possibly in the 1,000 to 10,000 range. But there are approaches that aren't, such as the one that Sawchuk and Strand have been looking at.

C. I think that any of the multiplexing techniques are going to be fundamentally limited to small images, as you get a phenomenal matrix sort of representation with these schemes. I guess the ideal case would be if you could process continuous data, then the space-variant processor would have a space-bandwidth product equivalent to a regular space-invariant processor.

C. It's been mentioned that a digital processor is much better than a 1-D optical processor in that the digital equivalent is just a matrix-vector multiplication, which is a quick, easy thing to do. But it's not that easy to do, say, a thousand component vector times a 1000 x 1000 matrix - it takes several milliseconds, it takes a rather big box, and it's power consuming. The optical system has a potential of much faster times, say, several microseconds.

C. Right now there's about four orders of magnitude difference between the speed of what I believe to be a buildable incoherent matrix-vector multiplier with reasonably small matrices, say 100 x 100, as compared to a floating point system array processor. There are military systems that are faster than a commercial floating point processor, but not four orders of magnitude faster. And power consumption is something that is extremely important and it is something we've never addressed in this community when making comparisons with competing technologies.

C. Maybe we should address the question of why these things haven't been packaged in the past. At a conference in 1972 on future directions in optical processing, people said things about packaging and how are we going to make these things viable, and they said a lot about spatial light modulators. Well, I see in the 1-D processing case you have a wide range of fairly available devices.

R. Who said they were going to spend money on packaging? Nobody did. Everybody said that what we really need is input-output, but nobody was packaging. So somehow you have to get the private sector or the military so involved that they want to make these things usable.

C. There is a problem in digital processing that apparently has not been solved. For image restoration, image enhancement, and the like in which there is a man in the loop, how do you know the image is restored or that you have some representation of the image back? How does this impact optical processing?

C. What I think you are getting at is that with a digital system you can re-program it very easily. If you have a wide range of software available, you can interact with the system in real time. But with optics it has been difficult to adjust the system and have a human interaction in the loop. With various real time devices that are coming out, that might be easier, but it's a difficult problem for optics and for digital systems, too.

C. Getting back to Harper's question, there's a great need to process images with a large number of points. 5,000 x 5,000 is not at all unrealistic, especially for earth resources pictures, radiographic images, and the like. We were talking before about examples of where you would like to do space-variant restorations. In almost any kind of radiographic image you have space-variant degradation of some sort, and if you're interested in details there you'd like to do restoration. Also, we ought to think a little more about discrete time and discrete space systems because digital systems operate this way. If we're going to build some optics, we'd better make it interfaceable with a digital system. I think things like the matrix-vector description of a space-variant processor are important and should be looked at a good deal in the future.

C. We have to realize that packaging any coherent system is very environmentally difficult, especially one that has any kind of interferometric detection. I think people in research view that as more of a development task, they like to come with the ideas but they wish someone else would worry about putting it in a little box. An impressive packaging success story is the laser gyroscope which is going into the Boeing 767.

C. There is another application area besides image processing, and that is the evaluation of integral transforms that are space-variant.

C. If we're setting the tone for the whole discussion of where to go optically, I believe we have to consider digital systems in the sense that they provide the control and decision making functions. When we go out of basic research and start talking about applications, we're trying to take the man out of the loop

somewhere. Bob talks about real time, you have to take the man out of the loop in digital systems. With that in mind, we can think of optics as providing partial solutions for problems in the sense that it is a preprocessing function, and then let the digital system work. That makes some of the basic research look attractive.

Processing flexibility by hybrid optical/digital techniques

David Casasent and B. V. K. Vijaya Kumar

Carnegie-Mellon University
Department of Electrical Engineering
Pittsburgh, Pennsylvania 15213

Abstract

Achieving flexibility in optical pattern recognition systems by novel hybrid optical/digital techniques is considered. Increased use of digital pre- and post-processing together with synthetic filters or discriminants and pattern recognition techniques beyond matched filtering are discussed.

1. Introduction

Pattern recognition (PR) [1-3] is one of the most complex and all encompassing data processing problems that we face in the 1980s. Sensor technology [4-5] has recently achieved the capability to produce data of sufficient quality and at such high resolution and data rates that exceed our ability to process and analyze it in a timely manner [6-7]. Optical processing has long been attractive for its real-time and parallel processing features. However, the flexibility and repertoire of operations achievable in optical processors remains quite limited and hence their use in practical PR systems has not yet materialized.

The optics community has recognized this shortcoming and has recently responded with a variety of new algorithms and architectures. Many of these concepts use ideas that are well-known in the mathematical and digital pattern recognition literature. In this paper, we will review these advanced PR approaches with attention to how they increase the processing flexibility of optical systems. The central unifying theme chosen is hybrid optical/digital processing, whereby the advantages of optical and digital technology can best be combined. Although PR is the specific application area discussed, the concepts and philosophies advanced are of use in other areas as well.

The accuracy of an optical processor is a second major area of concern. It is not directly addressed in this paper, but is of concern. Thus we note that optical numerical processors [8-9] using residue arithmetic and operating on digital data are also the subject of current research. Of more direct concern (to the image PR topic of this paper and the above issue of accuracy) is the concept of data SNR and system dynamic range. It has recently been shown that in most real-world cases detection performance is data-limited rather than processor-limited [10]. Thus the superior precision obtainable in a digital system is of less concern. Basically, the point reduces to whether the dynamic range of the optical system exceeds the SNR of the input data. If this is the case (as it appears to be), the optical system introduces no errors in the processing and its accuracy is thus adequate. Thus, the major concern of this paper is the flexibility of the optical system and therein the hybrid optical/digital processor concept and algorithm flexibility issues.

2. Problem definition and solution conceptualization

The PR problem can be described as determining the presence and location of a stored key object (reference image) in an on-line sensed image. The key issue in all practical PR problems is to maintain recognition when the input and reference differ for various expected and practical reasons. The two major classes of image differences of concern can be described as geometrical and textural. Geometrical differences between the input and reference images arise due to scale, rotation, aspect and other such errors. Textural differences are due to seasonal or other temporal differences and to multi-sensor effects (e.g. the reference and input imagery are taken at different times and/or from different sensors).

The concept of an observation space on which to operate is the first issue of concern. Fourier transform (FT) and image planes are the most obvious and most easily realized choices (Sect. 3). Statistical PR issues are given far more attention in digital and mathematical formulations than in optical systems. This trend should be altered. Use of digital processors for simulation and control of optical systems and electronic techniques for data normalization, detection and estimation also deserve note (Sect. 4) as do alternate PR methods beyond classical matched spatial filters (MSFs) (see Sects. 6-7).

In Fig. 1, the general block diagram of a flexible and practical hybrid optical/digital PR system is shown. A discussion of this schematic is of use in providing an overall introduction to the subjects of processing flexibility, practical PR problems, and hybrid optical/digital PR systems. From this figure, we notice two major elements: the digital pre- and post-processor. We also note the use of synthetic references or discriminant functions rather than conventional MSFs. In most practical PR cases, the input data must be preprocessed before being input to the optical processor. Specific examples of this on-line pre-processing are noted in Sect. 5. In many cases, appropriate synthetic reference functions (or more appropriately: filters or discriminant features) can be generated off-line by digital techniques. We distinguish between deterministic (Sect. 6) and statistical (Sect. 7) versions of these synthetic discriminants. We also note that a discussion of such issues is quite applications oriented and thus our discussions of such systems will refer to

specific PR problems to provide the necessary in-depth detail required to select: observation spaces, pre-processing techniques, and synthetic discriminant functions. In Sects. 4 and 8, we briefly address the issues of digital post-processing. However, the major theme in Sects. 8 and 9 is the use of advanced PR techniques beyond the MSF and the fact that such systems can be realized more efficiently by hybrid optical/digital architectures.

3. Image and FT plane sampling

The hybrid optical/digital PR concept shown in Fig. 2 is adapted from descriptions of a system under development at the Engineer Topographic Laboratory (ETL) [11]. This system is simple in concept and yet serves to introduce many key hybrid optical/digital PR concepts. The input image is scanned and the FT and image of 625 sampled scene regions are formed sequentially at both P_{1F} and P_{1I}. A wedge/ring detector [12] placed at P_{1F} provides directional and spatial frequency information on each sampled scene region. The output from a 32 x 32 element solid-state detector at P_{1I} is used to obtain statistical data such as the mean and variance for each image region.

These wedge, ring and statistical data, plus context information, are analyzed in a digital post-processor to determine input image content. Initial tests of this concept have proven useful in locating structured road regions of a scene [11] and in the analysis of the textural information content of radar and visible imagery [12]. Our major concern with this concept is its use of digital post-processing on multi-domain optically generated data and its use of statistical, textural, and context image data. These latter aspects of imagery have rarely been used in prior optical processors. Initial tests have indicated that significant image information is contained in such data [11, 12].

4. Digital/electronic adjuncts

Digital techniques have extensive use in the simulation [13] and control [14] of optical processors, plus in the production of computer generated holograms [15]. The conversion of sensor data to optical transparencies requires electronic interfacing. These uses of digital and electronic techniques are of less concern in this paper and we will thus concentrate on cases in which the digital and/or electronic system performs a more major processi role. Recent advances in smart sensors [4-5] and solid state detectors have made digital/electronic pre-/post-processing feasible. Advanced optical PR systems should capitalize on such advances in related disciplines.

A recent in-depth case study of optical word recognition on a microfilm data base [16] has resulted in the hybrid optical/digital system architecture shown in Fig. 3. This system has three features of concern in this present paper. First the dc value of the FT of the input page is a useful measure of the data content of the input text and can be used to control an input attenuator (A) to provide data normalization, thus simplifying the output detector decision system. Second, this particular application contained three classes of data, which were found to be possible to separate into classes by simple FT plane analysis. Third, three detection criteria (threshold, area and volume detection) were analyzed and the more advanced methods were found to be of considerable use and to be easily realized by simple detector modifications. Such techniques require more extensive data bases to determine their usefulness.

5. Real-time digital pre-processing

In this section, we consider several pre-processing operations that need be performed on the input and reference data to facilitate correlation and to increase the system's flexibility. These operations are best seen by considering specific PR case studies.

A major source of error in PR systems is the presence of geometrical differences between the input and reference scenes. By correlating coordinate transformed versions of the data, a PR system invariant to different geometrical distortions between the input and references images can be produced. The coordinate transformation chosen determines the geometrical error to which the processor is invariant. The general system block diagram of Fig. 4 is quite simple. It is a conventional frequency plane correlator [17] with a coordinate transformation pre-processing box. The combination of a coordinate transformation pre-processor and a space-invariant optical correlator results in a space-variant system, whose flexibility and practicality exceed that of the normal system.

The CT operation can be achieved in a digital pre-processor and since the operation must be performed in real-time on the input data, this class of system is treated in this present section. The details of this type of processor and demonstrations of its use in many cases have been described elsewhere [18]. One example of a rotation-invariant space-variant system is shown in Fig. 5. The inputs (top) are diatom cells in different orientations. The optically produced correlations are shown in the bottom and their cross-sectional scans in the center of Fig. 5. As seen, the system is invariant to rotational differences between the input and reference and the rotational differences can be found from the location of the correlation peak.

As a second case-study for which on-line image pre-processing is needed, we consider the correlation of a sensed input image and a synthetic reference image. This scenario arises in an advanced cruise missile terminal guidance problem. Figures 6a and 6b show typical examples of the sensed and synthetic data. The low output optical correlation obtained is shown in Fig. 6c. This system performance is due to: the considerable image texture present in the sensed scene and not in the synthetic one; the added small-surface detail (e.g. the roof of the building) present in the sensed image; the presence of objects (e.g. cars and trees) in one

scene and not in the other; contrast reversals and gray level differences in the scenes, etc.

These differences constitute noise rather than information, since they are not common to both scenes. With the help of Technology Services Corporation, we have considered application of various pre-processing operators to such scenes. Those digital pre-processing operations considered include: histogram equalization, gray level modification, edge enhancement, and small surface lysing. The first two operators make large image areas more uniform in intensity. The third operator enhances edges and structural scene data, whereas the fourth operated removes small surfaces and objects present in only one of the imagery.

The results of such operators are shown in Figs. 7a and 7b. They result in a considerably enhanced correlation peak SNR as seen in Fig. 7c. These pre-processing operators must again be realized in real-time and hence are included in our present discussion. Advances in smart sensor technology as well as digital and CCD hardware make such operations possible in conventional technology. We feel that optical processors should make use of such advances in other technologies by altering system architectures and algorithms as shown, rather than by striving to achieve such operations within the optical system itself.

6. Synthetic discriminant functions (deterministic)

In the third class of hybrid optical/digital processors that we consider, the digital computer is used to synthesize the reference function(s). These reference functions are synthetic, being produced from linear combinations of mathematical functions. The weighting coefficients and basis functions used are determined after extensive covariance matrix analysis and testing. Because of this, the operations required to synthesize the optimum reference filter must clearly be performed by a digital processor. Moreso, since the same reference filter(s) can be used to recognize any input of a given class, the filter synthesis can be off-line. Hence, the title chosen for this class of hybrid optical/digital processors adequately reflects the new role played by the digital system.

Related efforts in this area have recently been reported. It was first proposed [19] to use N reference functions to recognize N patterns. When one pattern was present at the input, N correlation outputs resulted. A weighted linear sum of these N outputs could then be used to determine the input pattern present.

In more recent work, use of a nonredundant set of synthetic references was considered [20]. To recognize N patterns, only $K = \log_2 N$ references are used. The K output correlations are thresholded and constitute a K-bit output binary word that can (after decoding) denote the input present. Use of the linear discriminant function theory that is widely used in digital PR has recently been suggested to produce generalized matched spatial filters [21] for use in PR. Synthetic references that comprise the key features of biological algae have also been successfully used [22].

Yet another different but related approach [23] that provided more detailed description has been reported. This technique used linear basis function expansion and covariance matrix computation and diagonalization to determine a single multi-variant filter that could recognize all representations of an input object, with both intensity and geometrical differences present. In this approach, each input image g in the data set $\{g_n\}$ of imagery to be recognized is first written as a linear sum of basis functions ϕ_j, $g_n = \sum_j a_{nj}\phi_j$. The single desired reference or discriminant function h is likewise expressed as $h = \sum_j b_j\phi_j$. To determine ϕ_j and b_j, the cross-correlation matrix $R_{ij} = g_i \circledast g_j$ is computed. Diagonalization of R_{ij} yields the a_{nj} and ϕ_j. We then require the correlation $R_n = g_n \circledast h$ of any g_n with h to yield the same constant value c. If the ϕ_j are orthonormal, $R_n = \sum_j a_{nj}b_j = c$ and we can thus find b_j and hence h.

The data base chosen to demonstrate this concept consisted of infrared tank imagery of the same object in different aspect views. A portion of the training set $\{g_n\}$ used is shown in Fig. 8. Optical weighted MSF synthesis [24] was used to determine the optimum filter passband of spatial frequencies to use. This accounts for the intensity differences present in IR and multi-sensor imagery as discussed elsewhere [25]. The cross-correlation matrix R_{ij} was formed (Fig. 9) and diagonalized (Fig. 10), thereby determining the a_{nj} and ϕ_j values and functions. Requiring $R_n = c$ provided us with the b_{nj} values. The synthetic discriminant function h so produced is shown in Fig. 11. A transparency of it was formed. An optical matched spatial filter of it was then produced and placed in the Fourier plane of a frequency plane correlator. The cross-sectional scans through one of the resultant R_n correlations are shown in Fig. 12. All output correlations R_n of seven widely different aspects g_n of the tank yielded quite identical peak R_n values as predicted.

This philosophy in which digital processing techniques are used to compute synthetic discriminants (by deterministic methods) off-line represents a major new hybrid opticla/digital pattern recognition approach that merits far more attention.

7. Alternate feature spaces and hybrid PR techniques

Within the PR and digital image processing literature, many alternate PR techniques exists beyond the simple matched spatial filter. Hybrid optical PR algorithms and architectures should make use of such available research.

As one example of such a hybrid PR system, we consider the use of invariant moments for PR. It has been shown that linear combinations of the ordinary moments (up to third-order) of a scene can be combined to realize seven invariant-moments [26]. Digital experiments have shown [26-28] that these moments are invariant

to all geometrical and intensity differences in an image.

Computation of the ordinary moments m_{pq} of a high-resolution 2-D scene is quite complex. However, an optical system such as the one in Fig. 13 can easily compute the m_{pq} for a given f(x,y) input when appropriate masks are used at P_2. When g = 1, x, y, xy, x^2 the P_3 output u_3 equals m_{00}, m_{10}, m_{01}, m_{11}, m_{20}, respectively, etc. A dedicated digital post-processor can then compute the invariant moments ϕ_n from these m_{pq} values.

This particular hybrid PR concept [29] and other such ones merit research analysis. The system architecture in Fig. 13 demonstrates many specific features of hybrid processing. Since the dynamic range of the m_{pq} are 30 dB whereas the ϕ_n are quite large (200 dB) and since the operations required to obtain the ϕ_n from the m_{pq} involve addition, subtraction, multiplication and division, the hybrid architecture shown appears to be well chosen.

Such a proper assignment of operations to the optical and digital processing portions of a system is essential. Likewise, this selection should be made with speed and component requirements in mind. In instances such as the one shown, digital processors can play a considerable role as post-processors.

8. Summary and conclusions

In prior sections, we have considered five classes of hybrid optical digital systems. In each case the role of the digital processor differs; but, in all cases, the combination of an optical and digital processor is chosen to increase flexibility.

In the cases considered in Sect. 4, digital and electronic systems were used as adjuncts to optical systems. Cases included were: digital simulation, digital control and data normalization plus alternate detection methods using advances in solid-state detectors and electronics.

In the second class of hybrid system (Sect. 3), the digital system was used as a post-processor operating on area sampled image and frequency plane data. This system is of particular interest as it is one of the few optical processors to employ statistical data analysis.

The third class of hybrid system (Sect. 5) employed digital pre-processing. The specific digital pre-processing techniques included must be realized in real-time on the input imagery. They include: histogram equalization, gray level modification, edge enhancement, and small surface lysing, plus coordinate transformations.

The fourth class of hybrid system (Sect. 6) also used pre-processing by the digital system. However, in this case, extensive off-line processing was used to produce synthetic discriminant functions for pattern recognition.

Such departures from classic matched spatial filtering are essential. An example of such alternate pattern recognition methods (our fifth class of hybrid system) was presented in Sect. 7.

The invariant moment technique advanced in Sect. 7 is typical of the novel hybrid PR systems that deserve further attention. The optical community must become more aware of digital PR techniques, smart sensors, and digital/electronic hardware advances. Such items must be incorporated into future hybrid systems if optical processors are to become practical. More attention to statistical techniques such as in [11] and [30] should be included in optical processors. Optical researchers should also take care to insure that they address real and practical problems for which digital methods are not adequate.

9. Acknowledgments

The partial support of the Air Force Office of Scientific Research and the National Science Foundation for much of the work reported upon is gratefully acknowledged. The authors thank all of the present and prior members of the Optical Image Processing Laboratory at Carnegie-Mellon University whose work we have summarized.

10. References

1. W. Pratt, Digital Image Processing, Wiley, New York (1978).
2. R. C. Gonzalez and P. Wintz, Digital Image Processing, Addison-Wesley, Reading, Massachusetts (1977).
3. K. S. Fu, editor, Syntatic Pattern Recognition Applications, Springer-Verlag, New York (1977).
4. Smart Sensors, SPIE Proceedings, Vol. 178 (1979).
5. Advances in Focal Plane Technology, SPIE Proceedings, Vol. 217 (1980).
6. W. E. Silvertson, Jr., and R. G. Wilson, Optical Pattern Recognition, SPIE Proceedings, Vol. 201, p. 17 (1979).
7. R. G. Wilson, W. E. Silvertson and G. F. Bullock, IEEE Computer Society Conference on Pattern Recognition and Image Processing, August 1979.
8. Numerical Optical Processors, SPIE Proceedings, Vol. 185, No. 1-5 (1979).
9. Optical Numerical Processors, Proceedings of the International Optical Computing Conference, Vol. 232 (1980).
10. D. J. Granath and B. R. Hunt, Applied Optics, Vol. 18, p. 36 (1979).

11. R. D. Leighty and G. E. Lukes, Optical Pattern Recognition, SPIE Proceedings, Vol. 201, p. 27 (1979).
12. H. C. Kasdan, Optical Pattern Recognition, SPIE Proceedings, Vol. 201, p. 1 (1979).
13. M. Saverino, Master's Thesis, Carnegie-Mellon University, Pittsburgh (1978).
14. D. Casasent and W. Sterling, IEEE Transactions on Computers, Vol. C-24, p. 348 (1975).
15. A. W. Lohman and D. P. Paris, Applied Optics, Vol. 7, p. 651 (1968).
16. D. Casasent, F. Caimi and J. Hinds, Optical Engineering (to appear).
17. A. Van der Lugt,.IEEE Transactions on Information Theory, Vol. IT-10, p. 139 (1964).
18. D. Casasent and D. Psaltis, Proc. of IEEE, Vol. 65, p. 77 (1977).
19. H. J. Caulfield and W. T. Maloney, Applied Optics, Vol. 8, p. 2354 (1969).
20. B. Breunecker, R. Hauck and A. W. Lohman, Applied Optics, Vol. 18, p. 2746 (1979).
21. H. J. Caulfield and R. Haimes, Applied Optics, Vol. 19, p. 181 (1980).
22. H. Fujii and S. P. Almeida, Applied Optics, Vol. 18, p. 1659 (1979).
23. C. Hester and D. Casasent, Applied Optics, (to appear in June 1980).
24. D. Casasent and A. Furman, Applied Optics, Vol. 16, p. 1662 (1977).
25. D. Casasent and D. Munoz, Optical Pattern Recognition, SPIE Proceedings, Vol. 201, p. 58 (1979).
26. M. K. Hu, IRE Transactions on Information Theory, Vol. IT-8, p. 179 (1962).
27. R. Wong and E. L. Hall, Computer Graphics and Image Processing, Vol. 8, p. 16 (1978).
28. S. Maitra, Proceedings of IEEE, Vol. 67, p. 697 (1979).
29. D. Casasent and D. Psaltis, Optics Letters (to appear).
30. J. Duvernoy, Applied Optics, Vol. 18, p. 2737 (1979).

11. Figures

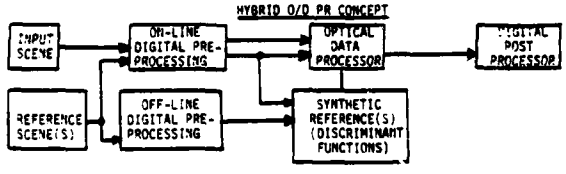


Figure 1 General block diagram of a hybrid optical/digital processor.

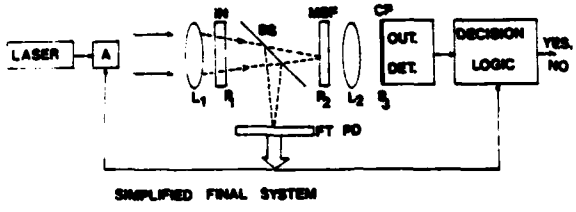


Figure 3 Hybrid processor for optical word recognition [16].

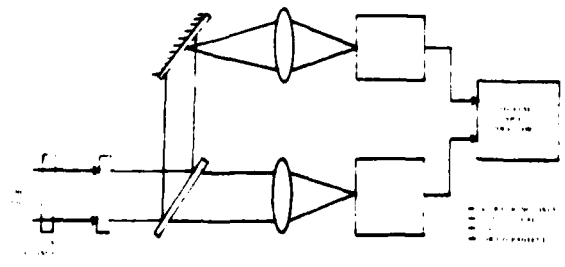


Figure 2 Schematic diagram of a hybrid processor using image and frequency plane sampling (adapted from [11]).

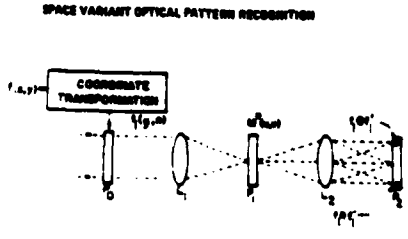


Figure 4 Space-variant hybrid processor.

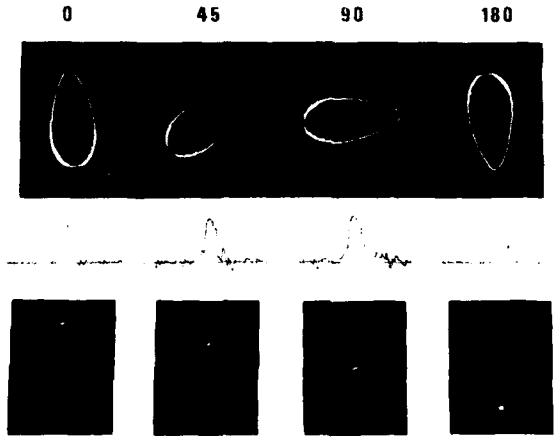
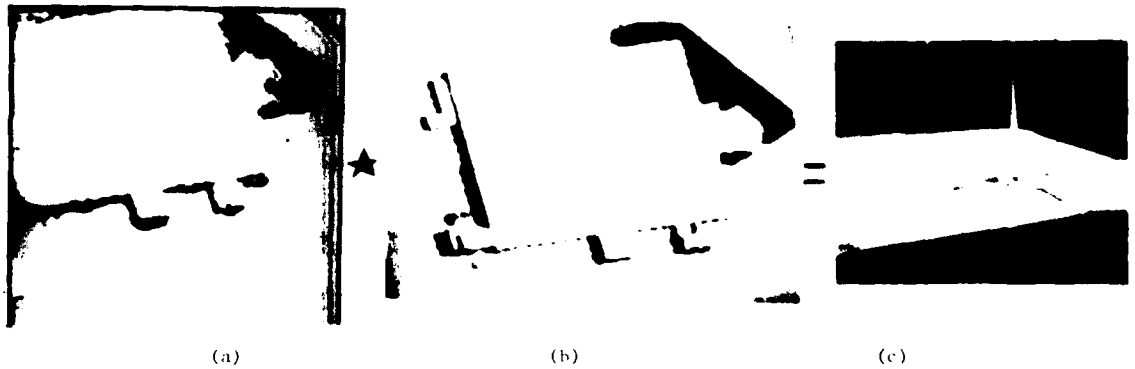


Figure 5 Input (top), output correlation (center) and cross-sectional scans of the correlations (bottom) from a space-variant rotation-invariant hybrid processor.

(a) (b) (c)

Figure 6 Input (a) and reference (b) images and with resultant output correlations (c) of a sensed input and synthetic reference.



(a) (b) (c)

Figure 7 Input (a) and reference (b) images and with resultant output correlations (c) of a sensed input and synthetic reference using real-time digital pre-processing.

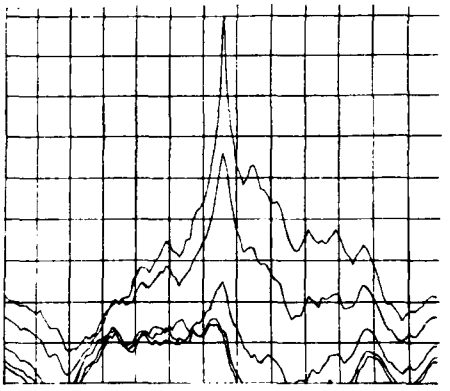
Figure 8 Representative imagery for synthetic discriminant feature processor [23].

Figure 11 Synthetic discriminant function generated [23].

CORRELATION MATRIX EXPERIMENTAL

ASPECT	1	2	3	4	5	6	7
1	2.42	0.29	0.15	-0.24	-0.09	-0.12	0.10
2	0.29	2.19	0.10	-0.14	-0.10	-0.07	0.06
3	0.15	0.10	4.95	0.02	-0.35	-0.30	0.03
4	-0.24	-0.14	0.02	1.51	0.04	-0.02	0.04
5	-0.09	-0.10	-0.35	0.04	3.87	0.03	0.02
6	-0.12	-0.07	-0.30	-0.02	0.03	3.93	0.02
7	0.10	0.06	0.03	0.04	0.02	0.02	0.22

Figure 9 Correlation matrix for seven images in Figure 8 set [23].



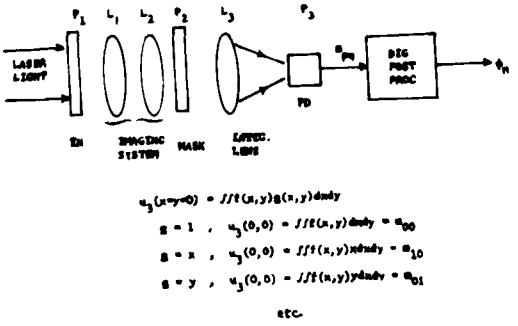
CROSS CORRELATION USING AVERAGED FILTER

Figure 12 Cross sectional scans through the correlation peak output from the correlation of Figures 8 and 11 [23].

	g_1	g_2	g_3	g_4	g_5	g_6	g_7
ϕ_1	0.64	-	-	-	-	-	-
ϕ_2	-0.08	0.68	-	-	-	-	-
ϕ_3	-0.02	-0.01	0.44	-	-	-	-
ϕ_4	0.07	0.04	-0.00	0.82	-	-	-
ϕ_5	0.01	0.02	0.03	-0.01	0.51	-	-
ϕ_6	0.02	0.01	0.02	0.01	-0.00	0.50	-
ϕ_7	-0.08	-0.06	-0.01	-0.07	-0.01	-0.01	2.1

GRAM-SCHMIDT BASIS

Figure 10 Diagonalized correlation matrix [23].



$$u_3(x=y=0) = \iint f(x,y)g(x,y)\,dx\,dy$$

$g = 1$, $u_3(0,0) = \iint f(x,y)\,dx\,dy = a_{00}$

$g = x$, $u_3(0,0) = \iint f(x,y)x\,dx\,dy = a_{10}$

$g = y$, $u_3(0,0) = \iint f(x,y)y\,dx\,dy = a_{01}$

etc.

Figure 13 Hybrid invariant moment processor [29].

Discussion (B. V. K. Vijaya Kumar; Discussion Leader: Nicholas George)

C. I want to point out that there is a growing interest and need for optical systems which do very limited tasks, i.e. spectral analysis or correlations, that can contribute to a larger system. In that case, packaging becomes an important consideration. One must build a system that is compatible with the larger system it is going to fit into, in terms of size, weight, power consumption, and input/output. A different way of looking at hybrid optical systems is that one builds a small package which fits into an overall processor.

Q. I feel that the weak link here might be things like A/D conversion. Will that slow down your ultimate process?

A. In the case of invariant moments, we are looking at only a small number of values at the interface, since computation is done in the processor.

C. I've heard references from several speakers to the engineering problems associated with putting optical components into systems. I think we need to be careful in the university research atmosphere of not putting too much emphasis on that as a major limitation. I think the engineering aspects are better addressed in the context of industrial research than in the context of university research.

Q. What type of question do you think the university researchers should be asking themselves?

A. The questions of function to be implemented, form and function, not fit.

R. We shouldn't be doing development in the university, but there may be concepts that might be found to allow us to package optical systems more efficiently and thus make optical processors more practical.

Q. Where does pattern recognition stand right now? Can we pretty well pattern recognize anything we want, or is it a wide open field?

A. In the medical community they have a fair degree of success, but trying to extract information out of natural scenes from real time sensors or photographic sensors is very difficult.

Q. Where could people who are working outside the military community get a set of pictures which they could use as a standard for working and sorting?

A. The problem of images to work on is very difficult. I [Harper Whitehouse] was associated with the TTCP, Tripartite Technical Cooperation Program, with the U.S., England, Canada, Australia, and New Zealand just with regard to I.R. images. It was agreed approximately a year ago that a data base would be established and would be made available on computer compatible tape. That study to determine the feasibility of generating a data base was assigned to J Advisory Group, number 11, which was formed about a year ago. I believe anybody who wishes to get information on scenes in the I.R. would be well advised to contact J Advisory Group through the Washington representatives.

C. A few years ago, we got interested in the issue of clutter and contrast in imagery. We found that the Air Force was interested in having pilots rank scenes as to the expected degree of difficulty in acquiring certain targets, and they were interested in the issue of what's a cluttered scene. They had pilots rank different aerial scenes according to clutter, and we were looking for some objective, mathematical measures for clutter. It makes me think that in some groups there have been varying degrees of gradings of aerial imagery, but I imagine it is a very task-dependent type of thing.

C. A point about compatibility: even in the digital case it hasn't been fully accomplished yet, there are very few instruments you can get and just plug together.

Q. Talking about hybrid schemes and optical processing in general, what does the group see as the total mission or goal? It seems that optical data processing is trying to compete with the Cray-1 computer, and that may not be a good place to work. We would rather identify specific applications like a pattern recognition system that handles a specific task, or a spectrum analyzer. You look at what you have to do, and you know that optics does some things, like Fourier transforms, very well. Do people see a mix of applications?

A. There was reference to this question in 1972 at the meeting at Carnegie-Mellon and it was interesting that there was really no one point where anyone gave the answer. I think we should recognize that what you are doing in your own labs that you are the most interested in could well be the direction that is moving ahead. These directions have to come out kind of subtly.

R. You can specify in the negative that you shouldn't be working in problems that can be done digitally.

C. I was intrigued by the use of invariants. Tom Newman at Texas Tech is doing very interesting work in digital pattern recognition based on a group theoretic approach, and the invariants you are calculating are different. It is an interesting approach because statistical pattern recognition has not provided all the answers that people though it would, and we need some novel approaches.

Comments by Tom Newman: My feeling about invariance is that there are so many effects that are actually deterministic in nature rather than stochastic that, if we are not careful, we can get into the statistical realm a bit too soon. The mathematical theory of invariance is quite old and to a certain extent a well developed area, an area which I feel is highly appropriate to be looked at. In John Walkup's talk, the definition he gave for space-variance and space-invariance would be given in a group theoretical context as "Spatially variant systems have a group of invariance which is the translation group." In many other systems it is equally possible to identify a particular class of transformations which lead to the images or objects under some invariance. Consider the Fourier transform, you get something there which is not classical invariance but is related to it in that the translation becomes a phase change, and this is also dealt with in invariance theory. In my own work I am not really in optical processing per se, but I look at motion as it appears in images and how we can handle problems associated with motion. Not only can you handle the motion, and many other things permitting the same general model, but also certain things like rotation and change of perspective. Most recently I've been concerned with what happens when you take 3-D images and project them onto a 2-D image plane. By examining data in the image plane, we seek to determine what is happening in the 3-D world. I think this is a very important problem. We have to realize that in the front end of a 2-D optical processor there was an ultimate recognition problem which doesn't really lend itself well to optical processing.

Nonlinear optical processing

A.A. Sawchuk and T.C. Strand

Image Processing Institute, Department of Electrical Engineering
University of Southern California, Los Angeles, California 90007 USA

Abstract

A review of the state-of-the-art in nonlinear optical information processing is presented. Due to space limitations, the empnasis is on nonlinear point functions. The three major techniques of point nonlinear processing are: halftoning; intensity-to-spatial frequency conversion; and direct nonlinear processing using the inherent characteristics of image detection and modulation systems. Application examples and real-time implementation of these techniques are described. Other less well-developed methods of nonlinear processing are covered briefly and references to all these techniques are given. Some prospects for future research are also presented.

1. Introduction

There is a great general need for systems that can perform fast, parallel multi-dimensional operations on signals with large time-bandwidth and space-bandwidth products. The time-bandwidth or space-bandwidth product of a signal is a measure of its complexity, and in many cases, traditional analog or digital electronic systems are overburdened or simply inadequate. The need for great processing capability arises in guidance, control, image processing [1,2], radar signal processing [3], image pattern recognition [4,5] and machine perception.

The parallel nature of optical systems and their inherently large space-bandwidth product has led to the development of many systems and techniques for optical information processing [6]. A fundamental difficulty with optical processing has been the limited range of operational "software" available [7,8]. Thus, general nonlinear operations such as logarithms, power laws, and limiters have been very hard to implement, while linear operations such as correlation, convolution, and Fourier filtering have been relatively easy. Many new techniques of signal processing and pattern recognition require nonlinear functions as part of their operation, and these functions have been achieved digitally, although in serial form [2]. In this paper we describe a large number of techniques for achieving these and many other nonlinear operations in optical systems. Many of the techniques utilize real-time optical input transducers which can convert electronic or image information into a form suitable for input to an optical processor [9]. With most of these techniques, the processing is performed almost exclusively in an analog fashion. Recently, several concepts for binary [10-21] or residue [22-28] numerical optical processing have been developed in which the signals are processed as discrete levels within the system. This new approach holds much promise for the future if real-time processing speed, accuracy, and flexibility can be maintained. We show in this paper how some of these operations can be treated as a form of nonlinearity.

Since the class of nonlinear operations is so broad and ill-defined, we restrict discussion in this paper to the well-defined subset of point (memoryless) nonlinearities where the output at any point is a nonlinear function of exactly one input point. Although this excludes much research in nonlinear processing, it includes a significant fraction of the work done thus far and lays a groundwork for much of the work with nonlinear systems which combine a point nonlinearity with other processing. A broader range of nonlinear systems are discussed in [29] from which this paper has been abstracted. One important class of nonlinear processing systems which are not discussed here are those using feedback to alter the system transfer function in a nonlinear way. Many excellent review articles have been written on this subject and the interested reader is referred to them [30-34]. Several other reviews of nonlinear processing are also in the literature [32-35].

The emphasis in the description of systems and experimental results in this paper is on work performed by the authors and their colleagues. The reason for this biased emphasis is that this is the work we feel most qualified to discuss and the only work we feel qualified to discuss in detail. Although by no means exhaustive, the references cited in the text along with the brief discussions of the cited works should give a more general view of the breadth and depth of the research in the area of nonlinear optical processing.

The next several sections describe specific systems which directly perform point nonlinear operations on two-dimensional input functions. There are three major categories of techniques: 1) Halftone processing; 2) Intensity-to-spatial frequency conversion; and 3) Direct nonlinear processing. The best developed technique is halftone processing, and this section describes the principle and gives experimental results and generalizations such as pseudocolor. The next section describes theta modulation and a relatively recent nonlinear processing method relying on variable-grating mode liquid crystal devices. This technique uses the intensity-to-spatial frequency conversion characteristics of these devices to implement nonlinear functions. The last section describes electro-optical systems which directly implement nonlinear functions by utilizing inherent characteristics of devices.

2. Halftone processing

The halftone method of nonlinear processing is a two step procedure. The first step converts the continuous level two-dimensional input signal into a pulse-width modulated (ideally) binary input. This operation is exactly the halftone procedure used in the graphic arts to represent an image containing gray tones as a binary picture. The halftoning step uses a mask transparency called a halftone screen along with a high contrast (ideally, a sharp threshold) photographic material or real-time coherent optical input transducer.

With both photographic or real-time halftoning, the second step of the process is Fourier filtering and recombination of diffraction components in a coherent optical system. The many variations of halftone screens and filtering procedures permit great flexibility in the nonlinear functions that can be achieved.

2.1 Halftone processing principles

The halftone process has been used for some time in the graphic arts industry to control density transfer characteristics in photo reproduction. Halftone screens generally consist of a one-dimensional or two-dimensional periodic array of identical continuous amplitude transmittance profiles, each varying continuously from opaque ($\underline{t}$=0 or D=∞) to transparent ($\underline{t}$=1 or D=0) [36-43]. Figure 1 is a diagram of the preprocessing steps for a one-dimensional input. $D_p(x)$ shown in Fig. 1(a) is the one-dimensional continuous input, plotted logarithmically as a photographic density. A typical halftone screen periodic density profile is shown as $D_s(x)$ in Fig. 1(b). In the preprocessing step, the continuous input and halftone screen are placed together and are photographically copied (by imaging or contact printing) onto a high contrast recording medium that ideally has a sharp threshold response as shown in Fig. 1(d). Figure 1(d) shows the ideal transmittance of a high contrast reversal photographic material vs. an input scale D_{in}. D_{in} is proportional to the logarithm of exposure through the sandwich of input and halftone screen. The D_t shown in Figs. 1(c) and 1(d) is a clip level and is the maximum density through which the recording medium can be exposed. The clip level is logarithmically proportional to the controllable uniform illumination used; thus the effect of varying the illumination is to move the D_t value up and down on the axis of Fig. 1(c).

The transmission of the recording medium is ideally either 1 or 0 because of its high contrast characteristics. The halftoned version of a continuous density distribution such as in Fig. 1(a) will appear as shown in Fig. 1(e). All values of x for which the density is less than the clip level turn black and hence transmit no light. All values of x for which the density is greater than the clip level do expose the medium, resulting in unity transmission.

It is this halftoned picture, as shown in profile in Fig. 1(e), which is capable of yielding a nonlinearity when placed in a coherent optical system as shown in Fig. 2. If this halftoned input picture is made with a two-dimensional halftone screen, the Fourier transform plane is a two-dimensional array of points of light which will ideally be the centers of distinct spectral islands, each of which contains complete image information. This assumes that the picture is sampled at a adequately high rate. If the sampling rate is not high enough, the spectral islands will not be separable, and aliasing will occur. One of these diffraction orders is selected by a spatial filter and retransformed by the second lens to yield the demodulated output. The output will be a nonlinear version of the original picture, where the nonlinearity depends on the halftone screen and the diffraction order chosen.

Several references [39-41] consider in detail the dependence of the system transfer function on the halftone screen and diffraction order used. Input-output relationships are established as a function of system parameters, and many numerical examples for useful point nonlinearities are given. Some experimental results for these nonlinear functions are given in Sect. 2.2. Additional variations in the halftone screens are possible; using

nonmonotonic cell profile functions f(x) allows a nonlinearity with many slope changes to be achieved in the first diffraction order [42]. Related design algorithms have also been given by Matsumoto and Liu [43].

The copy medium used in the pulse-width modulation step of nonlinear halftone processing must ideally have a sharp thresnold characteristic. Although some photographic materials closely approach this ideal, tney have the disadvantages of slow, clumsy operation. Recently, many different types of real-time optical input modulators have appeared which can convert electronic or image information into a form for input to a coherent optical processing system [9,44-47]. Many different technologies are utilized in these devices, but at present, most of them have the smooth, linear transfer characteristics of photographic film used for ordinary continuous-tone applications. Devices which operate at television rates (on the order of 30 frames per second) and perform sharp thresholding remain unavailable. Details of some of these devices and experimental results for nonlinear processing are given later.

A general analysis of the halftone process can be used to predict the effects of the copying medium given a screen profile and diffraction order. Several analyses of the problem have used computer simulation to predict degradation of the input-output curves from the ideal. The results of this work are extensive [41,48]; only a sample is given here.

For smooth monotonic nonlinear functions such as the logarithm, the main source of output degradation is the linear section of the $\underline{t}$ vs. D_{in} characteristic curve as shown in Fig. 3. The saturation regions where the slope of the curve is zero for low and high values of D_{in} have less effect on the performance. Figure 4 shows these effects. Figure 4(a) is the ideal two decade logarithmic response, where the horizontal axis is plotted on a normalized scale. Figure 4(b) shows the degraded response for a copying medium having a photographic gamma (slope of the linear part of the density vs. log exposure curve) of 3.0. The curve showing the degradation was computed using a piecewise linear model for the curve g(.) of the copy medium, although any particular measured response curve can be used. In Fig. 4, the I_{out} response tends to fall below ideal for high values of normalized I_{in}.

For nonlinearities with sharp slope changes and sharp corners, such as the level slice function, the sharp rising and falling transitions have a reduced slope, and the sharp corners tend to become rounded [41]. The reason for this is that sharp threshold functions completely rely on the thresholding characteristic of the copy medium to attain their sharp slope. This effect can be seen by comparing the ideal level slice shown in Fig. 5(a) with the degraded results shown in Fig. 5(b). Given a copy medium with a finite gamma greater than one, it is possible to increase the effective gamma by making a copy of the first halftoned image. The overall gamma of the process will increase and the threshold will be sharper. However, this procedure is clumsy and impractical for real-time implementation.

The results of computer simulation of compensation for a logarithmic function is shown in Fig. 4(c). In this simulation, 30 discrete points in the halftone screen density profile and screen density values from 0 to 2 are assumed. The optimized output curve is seen to approximate the ideal result in Fig. 4(a) with much less error than before. Similar results for a level slice function are shown in Fig. 5(c). Here the optimization procedure is successful in improving the fit to the ideal, but cannot increase the finite slope on the sharp transitions at the boundaries of the degraded response. Details of these procedures and many additional simulation results are given in the references [41,48].

2.2 Experimental results

This section describes experimental results on nonlinear processing by halftoning. The first section summarizes results done by traditional photographic methods; the second section describes recent real-time implementation with coherent input transducers.

2.2.1 Photographic implementation

The use of halftoning to modify gray level transfer characteristics dates back to the late 1800's with the development of the modern graphic arts and photographic industry. As early as 1893 Abbe [49] experimentally noted the diffraction orders present in the Fourier spectrum of a sampled image. More recently, Marquet [50] and later Marquet and Tsujiuchi [51] noted that demodulating a halftoned photographic image in a conerent optical system produced various monotonic and nonmonotonic nonlinearities depending on the diffraction order chosen in the Fourier transform plane. Similar experiments and the basic idea of halftoning for nonlinear processing were given by Delingat [52]. Other halftone experiments were reported by Pappu, Kumar and Mehta [53], and Roychoudhuri and Malacara

[54] explored constrast reversal with halftoning.

Kato and Goodman [36,37] experimentally found that a commercially available halftone screen performed a logarithmic nonlinearity over two log units of exposure. They demonstrated the use of the technique for the separation of multiplied signals by homomorphic filtering. In homomorphic filtering the logarithm of two signals which have been multiplied is taken first. This log operation converts the signals into additive form so that traditional linear filtering can be used. Homomorphic filtering is effective for separating signals with a multiplicative noise component. Kato and Goodman demonstrated its application in separating periodic multiplicative noise from an image as shown in Fig. 6. Part (a) of this figure shows a face and a grating multiplicatively combined. Part (b) shows the homomorphically filtered result in which a halftone screen has been used to convert the input into additively combined signals which can be easily separated with Fourier filtering. Figure 6(c) shows the result of pure linear filtering without the halftone process to remove the grating. They describe additional experimental results for suppressing speckle noise and removing the screen grid from radiographs. Dashiell and Sawchuk [40] also made logarithmic halftone screens on a plotting microdensitometer. These screens also gave a good approximation to a logarithmic response over two decades of input dynamic range. Dashiell and Sawchuk also experimentally performed exponentiation having an output dynamic range greater than three decades.

Dashiell and Sawchuk have experimentally demonstrated level slicing using the first Fourier diffraction order as described in references [40,55]. Their halftone screen was a low contrast copy of a Ronchi ruling. Figure 7 shows their experimental results; part (a) is the original continuous tone scene and part (b) shows the isodensity contours of level slicing.

Much early work on halftone nonlinear processing was concerned with multiple isodensity or multiple isophote level slicing. Schwider and Burow [56] described a technique for filtering an image hologram to obtain isophotes. Delingat [57] showed that passing high diffraction orders of a halftoned image produced isophotes. Schneider [58] and Schneider, Fink and Van Der Piepen [59] described the design of special halftone screens and filtering systems to obtain isophote and isodensity contours. Liu, Goodman and Chan [60] have also demonstrated multiple isophote contours by filtering a halftoned image. Figure 8 taken from Strand [38] shows an example of multiple isophote processing. Figure 8(a) shows the input image and Fig. 8(b) shows three isophotes, each representing a doubling of the input amplitude.

Dashiell and Sawchuk [42] have demonstrated optical image quantization and intensity notch filtering by halftoning. They described a synthesis procedure for halftone cell profiles that are not monotonic. Using such a screen, they can achieve an arbitrary number of sign changes in the input-output curve using the first diffraction order.

Lohmann and Strand [61] and Liu [62] have performed analog-to-digital (A/D) conversion by halftoning using photographic film. Figure 9 from Ref. [61] shows a three bit A/D output. The results of the conversion appear serially as bit planes, each of which displays the information of a particular significant bit of the digitized image. The solid lines plotted on the I_{in} axis of the figure show the nonlinear transfer characteristic needed to produce the bit planes of the three-bit reflected binary or Gray code. The bottom curve showing I_{03} vs. I_{in} is the least significant output bit. The halftone process was used and the output results are shown scaled and superimposed on the drawing. I_{02}, the next most significant bit, has the same characteristic curve as I_{03} except that the horizontal axis is expanded by a factor of two. This expansion is achieved experimentally by attenuating the input by two and repeating the experiment. Similarly, I_{01} is the most significant bit and it is obtained by attenuating by a factor of four. In this way, each bit can be obtained sequentially by halftoning, and any analog input on the scale from 0 to 8 gives a unique three-bit quantized and digitized representation. Results of A/D conversion on a two-dimensional image are shown in Figs. 10 and 11 [63]. Figure 10 is the input image. Figure 11 shows the Gray code bit planes. The images on the left side of Fig. 11 were generated by the optical A/D conversion. Bit planes produced by a scanning microdensitometer are shown on the right side of Fig. 11 for comparison. The discrepancies near the perimeter of the image in the least significant bit are attributable to nonuniform illumination in the optical processor.

Liu [62] has designed a halftone screen which produces several bit planes with only one halftoned photograph. The different bit planes in decreasing order of significance are obtained by passing higher order diffraction terms in the Fourier filtering step. Liu [64] has also described another halftone procedure for A/D conversion which relies on an electronic logic array to produce the digital output at each image point. A photographic halftone procedure is used first as a quantizer to give a discrete input to the logic system.

2.2.2 Real-time implementation

There has been very little experimental work on real-time nonlinear halftone processing because of the lack of real-time optical input transducers with a threshold characteristic. Logarithmic and level slice nonlinear processing has been attempted with a standard Hughes liquid crystal light valve (LCLV) suitable for linear incoherent-to-coherent optical conversion.

The LCLV exists in both electronically and optically controlled versions. The optical LCLV serves as a real-time incoherent-to-coherent converter. Incoherent light impinges on a photoconductor, which in turn changes the local electric field across a liquid crystal layer. The change in electric field alters the local birefringence of the liquid crystal material. This birefringence pattern can be read out as a spatial amplitude modulation by placing an analyzer in the output beam oriented orthogonal to the initial polarization of the readout beam. The readout illumination can be spatially coherent or incoherent, but must be temporally narrowband. Additional variations in the device response are possible by introducing a twist in the alignment of the liquid crystal molecules during device assembly. The LCLV device operates at television frame rates (approximately 30 ms. cycle time) and generally is designed to have a linear response over two decades of dynamic range. Many references that discuss construction and operational details are available [44,45,47].

To obtain a real-time logarithmic nonlinearity [41], the characteristic curves of a Hughes 45° twisted nematic LCLV were measured. Figure 12(a) is the original characteristic curve of the LCLV, showing that the device has a smooth curve approximately that of photographic film with a gamma of 2 to 3. Using this data, an optimum compensated discrete halftone screen was made using the procedure outlined in Ref. [41]. This halftone screen had a fundamental spatial frequency of 3 cycles/mm, and was designed to work in the zero diffraction order.

The experimental setup used is shown in Fig. 13. The continuous level input is placed in contact with the halftone screen and imaged onto the control surface of the LCLV with incoherent illumination. The LCLV is read out with coherent light between crossed polarizers and a pinhole spatial filter is placed in the zero order in the Fourier plane of the system. Figure 12(b) shows the response curve of the system with the halftone screen in place. The result approximates a logarithmic transfer function with less than 5% error over one decade and less than 10% error over two decades.

To test the effectiveness of the logarithmic filtering system in another experiment, two crossed multiplicatively combined Ronchi rulings were used as an input picture for the experimental setup of Fig. 13. The period of these rulings was approximately 3 mm, much higher than the halftone screen period of 0.33 mm. The spectrum in the filter plane is shown in Fig. 14(a). Next, the logarithm halftone screen was placed in contact with the rulings. The filter plane spectrum is shown in Fig. 14(b). The difference in Fourier spectra between multiplicatively combined gratings and additively combined grating obtained by real-time logarithmic filtering is as follows: the additive spectrum components lie only on the x and y axis around the zeroth diffraction order in the frequency - domain while the multiplicative spectrum contains cross-term off-axis components. Figure 14 also shows the higher diffraction orders that arise due to the halftone screen. For simple logarithmic processing, these higher orders would be eliminated by spatial filtering.

The system has also been applied to an image filtering problem. For this experiment a picture with multiplicative noise was generated. The goal was to eliminate the multiplicative noise by homomorphic filtering; i.e., a sequence of a logarithmic transformation followed by linear filtering. Ideally, this would be followed by an exponential transformation but this is not essential in demonstrating noise reduction [41]. The exponentiation was not included in the following experiments. The noise generated for this experiment models that of a push-broom scanner that scans six lines at a time with six independent detectors. Any variation in the sensitivity or gain along the row of detectors gives rise to a periodic six-bar noise structure across the image. This effect is shown in Fig. 15(a). The density range of the image with the simulated scanner noise was 2.0D. This was chosen to match the 100:1 operating range of the halftone screen.

The noisy image was sandwiched with the halftone screen and imaged with incoherent illumination onto the LCLV. The resultant image was read out with a HeNe laser. A spatial filter was placed in the Fourier plane of the system to select a single (zeroth) diffraction order in the halftone spectrum. Since the screen effectively performs a logarithmic transformation, the pattern in the zeroth diffraction order of the Fourier plane consisted of the sum of the image diffraction pattern and the push-broom scanner

diffraction pattern. The scanner diffraction pattern consisted of a series of isolated diffraction orders which could be filtered out without significantly degrading the image. Thus in the output plane the image is reconstructed *without* the multiplicative noise. This is shown in Fig. 15(b). Without the logarithmic transformation, the noise and image spectra would have been convolved with one another making them impossible to separate.

These experiments demonstrate the feasibility of performing real-time nonlinear filtering for smooth functions using an LCLV or other real-time device. A real-time level slice experiment has also been performed with a standard linear 45° twisted nematic LCLV [65] and the results are comparable to the degraded level slice work shown in the simulations of Fig. 5(b). Until sharp threshold real-time devices become available, halftone nonlinear processing will be limited to smooth nonlinear functions.

2.3 Extensions and conclusions

Several extensions of nonlinear processing useful in image manipulations have been reported experimentally. Pseudocoloring is the process of associating colors with different gray levels in a monochrome image so the image is enhanced for display. One motivation for the use of pseudocolor is that the human eye is much more sensitive to small changes in color than to small changes in gray level. The pseudocoloring system described by Liu and Goodman [66] begins by halftoning the continuous tone input using a screen designed to produce multiple isophotes [60]. The halftoned image is then placed in a coherent processor as shown in Fig. 2 and is sequentially illuminated by red, green and blue monochromatic collimated plane waves. Different diffraction orders may be selected at each wavelength, and the resulting images for each color are summed photographically on color film at the output. The result contains different combinations of the three primary colors at each point as a function of input level.

Indebetouw [67] has described a pseudocoloring system which combines halftoning with elements of the theta modulation procedure discussed in Sect. 3. A one-dimensional halftone screen designed to produce a level slice is used in contact with the continuous tone input to expose a high contrast film as usual. This process is repeated with the exposure changed so that the effective position of the level slice on the I_{in} axis is shifted. In addition, the one-dimensional screen is rotated through a small angle between exposures. In effect, the various level slices are encoded as directional information on the binary copy film. The copy film is then placed in a coherent processor, except that a white light point source is used in front of the collimating lens to produce white light plane wave illumination. In the Fourier plane, level slice information recorded at a particular range of intensity values appears as a line of diffraction spots located at a unique angle. This pattern is modified with an angular array of color filters. After inverse transforming, the desampled image has various colors associated with different bands of gray levels. An advantage of this method is that the colored result is not formed sequentially and need not be summed photographically on color film.

The preceding principles and experimental results are a summary of the present state of halftone nonlinear processing. The procedure allows a large variety of nonlinear functions to be performed, but also has limitations. Two major limitations are that accurate halftone screens with sufficient gray level resolution must be available, and that a sharp threshold is needed on the copy medium (photographic film or real-time device) to ensure a binary input to the processor. The copy medium must also have a spatial resolution sufficient to record the finest detail in a sampled halftoned image of the original. Closely related to this is the tradeoff between spatial resolution in the copy medium and gray scale accuracy in the final output due to the pulse-width modulation nature of the process. In short, the overall performance of nonlinear halftone processing depends strongly on the binary, high resolution character of the copy medium. These properties can generally be achieved in photographic processes but are still unavailable in a real-time optical input transducer. A final limitation is noise in the form of speckle, interference fringes and other artifacts that appear in coherent systems. Careful precautions must be taken to avoid these effects. Finally, nonlinear processing using halftoning and incoherent spatial filtering is possible, in exactly the same way as the eye views a halftoned photograph. This procedure eliminates coherent noise problems, but is limited to monotonic nonlinear transformations of intensity. Achieving nonmonotonic functions requires the interference properties of coherent light.

3. Intensity-to-spatial frequency conversion

One very convenient method of obtaining point nonlinearities is through intensity-to-spatial frequency conversion. The idea is to encode each resolution element of an image with a grating structure where the period and/or the orientation of the grating is a function of the image intensity at the point in question. Assuming certain sampling requirements are met, each intensity level of interest is uniquely assigned to a different point in Fourier space and all points with a given intensity in the image are

assigned to the same point in Fourier space (assuming space-invariant operation is desired). Then a pure amplitude spatial filter can alter the relative intensity levels in an arbitrary way. The overall transformation can be found graphically as shown in Fig. 16 which shows an implementation of a level slice. In the figure, the quantity s is used as a generalized spatial frequency coordinate. In terms of polar spatial frequency coordinates, s typically varies either along an arc of constant radius ρ or along a radius of constant azimuth θ (see Fig. 17(b) and (c)). The first case is referred to as theta modulation and the latter case is referred to by the somewhat ambiguous term, frequency modulation. In both theta modulation and frequency modulation methods there is a certain connection between the number of intensity levels to be distinguisned, the bandwidth of the object, and the spatial frequency required for the modulated grating. For the general case of intensity-to-spatial frequency conversion, the spatial frequency is modulated along some arbitrary curve s in the spatial frequency plane (see Fig. 17(a)). A grating with a given fundamental frequency (ρ,θ) in polar coordinates, corresponding to a specific input intensity, produces a diffraction order at (ρ,θ) surrounded by a diffraction spot whose dimension is roughly equal to the bandwidth, B, of the object. If one wants to process N distinct intensity levels, then the curve s must be chosen so that N non-overlapping object spectra can be placed along s as indicated in Fig. 17. For theta modulation and for frequency modulation, this requirement sets a lower limit for ρ_0, the lowest fundamental frequency of the modulation. In fact, a rough calculation shows that ρ_0 must be on the order of N•B or larger, i.e., the lowest modulation frequency must be larger than the product of the image bandwidth and the number of gray levels. Obviously, if N is large, very high spatial frequency gratings will be required which is a limiting factor in these techniques. This would particularly be true if the optical system had to be capable of imaging these gratings. Fortunately, since spatial filtering of the grating spectra to select a single diffraction order is an integral part of the process, the space-bandwidth product requirements on the optical system are not so severe.

3.1 Non-real-time implementation

Of these two intensity-to-spatial frequency conversion techniques, theta modulation was the first to be discussed and demonstrated [68,69]. A practical problem associated with theta modulation as well as with frequency modulation is how to obtain the spatial frequency encoding. One solution to this problem for theta modulation which has been demonstrated involves the use of special halftone screens [59]. The screen profile is designed to produce halftone dots which have one edge whose orientation is a function of the intensity in the original image. The theta modulation technique is especially well suited to white-light image processing since the spectra at different orientations do not overlap as they would in general for a pure frequency modulation scheme. The pseudocolor system of Indebetouw [67] described in the previous section also uses halftone techniques to achieve theta modulation.

Lee has recently described a system which utilizes both theta modulation and frequency modulation [34]. A scanning interferometric system is used to encode each picture element of an image as a small grating segment whose frequency and orientation can be changed in accordance with input intensity levels. If real-time processing is not required, this system has the advantage of being able to more fully utilize the Fourier plane than either theta modulation or frequency modulation.

3.2 Real-time implementation

Frequency modulation has recently become a potentially viable means of achieving point nonlinearities in real-time with the advent of a device referred to as a variable-grating-mode (VGM) liquid crystal device [70,71]. The VGM device consists of a thin (<12μm) nematic liquid crystal layer sandwiched with a photoconductor between two electrodes (see Fig. 18). If a DC voltage is applied across the liquid crystal, a phase grating structure is produced in the liquid crystal material. As the voltage is varied the period of the grating structure is altered, as shown in Fig. 19. As seen in Fig. 19 the fundamental frequency of the grating structure can typically be varied between 200 and 600 lp/mm. Since the voltage across the liquid crystal can be locally controlled by projecting an image onto the photoconductor, the device functions as an intensity-to-spatial frequency converter. The attractive aspects of this device are that it can operate in real-time and it is a parallel processing device. Furthermore many useful nonlinearities can be implemented with very simple filters. For example, a level slice operation is obtained by placing a slit in the Fourier plane passing only a narrow spatial frequency band corresponding to a given input intensity range. The level slice can be continuously varied by merely translating the slit filter. The level slice function has been demonstrated with the VGM device [72].

The VGM device is still very much in the development stage. There are many problems to be solved before the device could be considered as a practical element of an optical processing system. Principal among the problems is one of defects in the grating

structure as seen in Fig. 20. These defects cause the diffraction orders of the grating to smear. As can be easily deduced from the earlier analysis, any smearing of the diffraction orders reduces the number of distinct intensity levels which can be processed. Therefore, it is necessary to keep the defects at a minimum if one wishes to process an image with a large number of intensity levels. Another important question which remains to be answered is the speed of the VGM device. Although current devices are slow and, due to the nature of the device, it is unlikely that high speeds could ever be attained, the goal of attaining speeds approaching television frame rates would provide an adequate processing rate due to the parallel operation of the device. Other areas that are being studied concern such issues as device lifetime, uniformity and reproducibility. It is expected that a better understanding of the underlying physics of the VGM effect will have a significant impact on all the problem areas discussed.

One application area of nonlinear processing which is of great general interest and which is of particular interest for work with current VGM devices is the implementation of optical logic [73]. Figure 21 shows how logic operations can be viewed as simple point nonlinearity operations. This figure assumes that two binary inputs are superimposed on a device and indicates the nonlinear response required to obtain various combinatorial logic functions. It is seen that the required nonlinearities are simple binary functions. These functions are particularly well-suited to VGM devices in principle since the desired nonlinearities can be obtained by simple slit apertures in the Fourier plane of the VGM device. The logic functions are also well-suited to the practical limitations of current devices because there are only a small number of intensity levels that need to be distinguished in a logic system. This in turn implies that defect-broadened diffraction orders can be tolerated.

Of course the interest in performing logic operations optically is again due to the processing advantage of a parallel processing system. Each resolution element of the processing system functions as an independent logic gate and all resolution elements are processed simultaneously. Thus even an inherently slow device can provide a high effective data processing rate. Furthermore, if the data to be processed are initially in the form of an image, having a parallel optical processor can obviate the need for scanning and serializing the data. Given an analog image, one still needs to perform an analog-to-digital (A/D) conversion, but that can also be done in parallel optically, either with a VGM device or with other nonlinear techniques discussed in this chapter [61,74].

Figure 22 shows how a VGM device could be used to perform two-input combinatorial logic functions. The two binary inputs are imaged onto a transparent photoconductor. This may be a completely incoherent imaging process. A spatially coherent readout beam is passed through the device. The read and write beams in this setup should be such that they can be separated either with a spatial filter or a color filter. A slit aperture placed in the Fourier plane is used to implement the desired logic operation. Experiments were conducted using two inputs which were chosen such that the output would be in the format of a truth table. The results are shown in Fig. 23. In these experiments the image area was approximately 1 cm^2. The images are written with a filtered Hg-arc lamp. An optical bias in the form of a uniform background illumination was included so that the zero input condition would correspond to a non-zero VGM frequency. This offset is necessary for implementing the NOT function or any other function requiring a non-zero output for a zero input. The image was read out with a collimated HeNe laser beam.

Not only can one implement the basic combinatorial logic functions with this system, but one can obtain any commutative Boolean function with a single pass through the system. Thus, for example, a full-adder can be implemented directly with a single device. Indeed since the VGM device produces a symmetric spectrum, one can filter the positive diffraction orders to obtain the sum bit of a full adder and simultaneously filter the negative orders to obtain the carry bit. The experimental setup for realizing a full-adder which simultaneously operates on all points of one image bit-plane is shown in Fig. 24.

By adding feedback to the system, it is possible to extend this technique to sequential logic. Furthermore, if a means can be devised of interconnecting the pixels on a given device, one can construct entire circuits on a single VGM device.

It should be noted in closing this discussion that even if the VGM device itself does not prove to be a practical optical processing component, the concept of having an element which performs a two-dimensional intensity-to-spatial frequency conversion in real time is so powerful that it warrants research into other possible means of achieving that function.

4. Direct nonlinearities

Direct nonlinear optical functions are achieved using the inherent nonlinear transfer characteristics of an optical recording medium or real-time image transducer. With this type of nonlinear processing, there is no pulse-width modulation, intensity-to-spatial frequency conversion or other type of intermediate mechanism. Thus these techniques offer the potential of simple systems that avoid the noise problems associated with many optical filtering techniques and have much less stringent space-bandwidth product requirements than systems which must modulate the input data.

4.1 Photographic implementation

Photographic film has been used since its invention to perform nonlinear image transformations. Although most photography uses the linear part of the density vs. log exposure curve to ensure linear recording, the limiting characteristics of the saturation region of the film curve and the nonlinear curve (toe) rising to the linear region have been used for artistic and scientific effect. Tai et al. [75] have obtained a direct logarithm by carefully choosing operating points on the characteristic curves of film. A two film procedure with a copying step is used, and a dynamic range of two decades is obtained without the need for a halftone screen. They have applied this procedure for homomorphic filtering of multiplied signals, with results comparable to those shown in Fig. 6.

Agfa [76] has produced a special photographic product called Agfacontour containing a sandwich of positive and negative emulsions. This combination gives the film a direct level slice characteristic which produces an isodensity contour output.

4.2 Real-time implementation

For real-time direct nonlinear processing, television systems and solid-state sensor arrays combined with electronic scanning and nonlinear processing circuitry have been used for nonlinear processing. The characteristic transfer curves can be electronically shaped so that the complete system, consisting of scanned image sensor, processing and display, performs a desired nonlinearity. Because of the scanned nature of these systems however, the parallel processing capabilities of optics are lost.

A real-time optical transducer which has been used for nonlinear processing is the Itek PROM (Pockels Read-Out Optical Modulator) [46]. This device is based on a $Bi_{12}SiO_{20}$ crystal which is both birefringent and photoconductive. The PROM is sensitive at blue wavelengths and is read out nondestructively at red wavelengths through crossed polarizers. By altering bias voltages in a particular sequence with the application of a control image, various linear operations such as contrast inversion and real-time spatial wavefront modulation can be done. Using a different control sequence produces an approximate real-time level slice effect. PROM devices are still in the research and development stage, but their flexibility in performing linear and nonlinear optical real-time processing makes them interesting candidates for future work.

The Hughes liquid-crystal light valve (LCLV) described earlier in Sect. 2.2.2 is another real-time device that has been applied to real-time direct linear and nonlinear processing [45,47]. The particular characteristic curves of such devices can be used for direct nonlinear processing in real time in exactly the same way as photographic film with the choice of positive or negative characteristic curves. An example of this is a real-time incoherent image subtractor which has been built using two LCLV's [77].

A direct implementation of A/D conversion in real-time using a special type of LCLV has been described by Armand et al. [74]. The system performs the sampling, quantization and digitization of a two-dimensional data array without scanning. The method uses the nonlinear device characteristics of a special LCLV having uniform parallel alignment of the liquid crystal material. In this configuration, the device exhibits a pure birefringent effect that varies with local electric field controlled through a photoconducting layer. The special LCLV is called a multiple period device; the overall relationship between the intensity transmittance of the device and the incident intensity at any point is given ideally by the sinusoidal curve shown with dashes in Fig. 25(a).

The digital results of A/D conversion at each image point may be output serially as a bit sequence or in parallel as bit planes like those shown in Fig. 11. The solid-line curves of Fig. 25 show the nonlinear transfer characteristic needed to produce the bit planes of the three-bit reflected binary or Gray code and their relationship to the dashed curves of sinusoidal device characteristics. When the output of Fig. 25(a) is thresholded at one half, a 1 output is produced above threshold and a 0 output below, as shown by the curves with solid lines. This thresholding can be done electronically following light detection by a parallel array of sensors. The threshold output in Fig. 25(a) is the least

significant bit of the three-bit Gray code. The other two bits are obtained by attenuating the input intensity to rescale effectively the horizontal axis. Use of the full dynamic range (0 to 8) gives the least significant bit. Attenuating the input by a factor of 1/2 (to the range 0 to 4) gives the first cycle of the characteristic curve shown in Fig. 25(b). The last (most significant) bit is obtained by using an attenuation of one fourth so the curves of Fig. 25(c) result. Note that any continuous input between 0 and 8 gives a unique quantized three-bit output. Although the outputs in Fig. 25 are the three bits of the Gray code, other A/D code conversions, such as the usual straight binary code, can be achieved by translating these curves left or right along the horizontal axis. This can be done by introducing phase-retardation plates with different delays along orthogonal axes into the crossed polarizer system.

The system can produce these bits in parallel by placing an array of three periodically repeated attenuating strips over the write surface of the liquid-crystal device, as shown schematically in Fig. 26. The strips have attenuation factors of 1, 1/2, and 1/4, and the image of the strips is in register with a parallel photodetector array with electronic thresholding in the output plane. All three bits are sensed in parallel in this way. The period of the strips should be much smaller than the inverse of the maximum spatial frequency of the input picture to avoid aliasing. A two-dimensional array of attenuating spots with a corresponding detector array can also be used instead of the linear strip array. Simpler but slower operation can be achieved by using only one detector array and sequentially uniformly attenuating the entire input array. Ideas such as this have been used for electro-optic A/D conversion, but none have used an optically controllable real-time device.

The experimental system shown in Fig. 27 was used to obtain the input-output characteristic of the pure birefringent LCLV and to demonstrate the concept. The incoherent source illuminating the input plane is a mercury-arc lamp. A fixed and rotatable polarizer pair in the input-light beam is used to vary the input-light intensity. The real-time device was a Hughes parallel aligned nematic LCLV with a CdS photoconductor. The short-wavelength cutoff filter eliminates wavelengths shorter than 493 nm to make sure that the write-beam wavelength is matched to the sensitivity range of the CdS photoconductor. The read light source is a xenon-arc lamp. Because of the dispersion of birefringence in the liquid-crystal material, the read light should have a narrow spectral bandwidth. An interference filter with a peak wavelength of 434.7 nm and a bandwidth of 18.4 nm was used to meet this requirement for the read light. With no picture in the input plane, the output intensity varies in a quasi-sinusoidal fashion with increasing input illumination because of the changing birefringence. If the amount of birefringence varied linearly as a function of the write-beam intensity, a strictly sinusoidal variation of the output intensity would be expected. However, a number of factors, including the optical nature of the liquid crystal and the photoconductor characteristic properties, affect the output curve and produce an approximately sinusoidal output whose frequency varies (monotonically) with input intensity. The experimental response curve obtained is shown in Fig. 28.

Although the theory behind the A/D conversion assumes a strictly periodic response characteristic, it is possible to produce the desired bit planes by using the quasi-periodic response curves of the actual device. The tradeoff is that nonuniform quantization results. The quantization levels obtained in this experiment are shown in Fig. 28.

There are no attenuating strips on the liquid-crystal device used in this experiment. Instead, the bit planes were generated serially. Also, the output was recorded on hard-clipping film rather than a thresholding detector array. A test target was generated that consisted of an eight-gray-level step tablet. The gray levels were chosen to match the quantization levels shown in Fig. 28. This test object was imaged onto the liquid-crystal device, and the output was photographically hard clipped to produce the least significant bit plane of a three-bit A/D conversion. Next the write illumination intensity was decreased, effectively rescaling the response curve of the device to generate the next bit plane. The last bit plane (most significant bit) was obtained by attenuating the write intensity again and photographing the output. The input and the three bit planes generated are shown in Fig. 29.

Although the output contains some noise, the experiment illustrates the principle of real-time parallel incoherent optical A/D conversion. It was found later that the computer-generated gray scale was somewhat noisy because of the grain of the high-contrast film used. It is possible that future experiments with clearer inputs and improved periodic light valves could produce better experimental results and more bits of quantization.

The potential A/D conversion rate can be estimated from typical parameters of currently available devices. The important parameters are device resolution (typically 40

cycles/mm), device size (typically 50 mm x 50 mm), and speed (generally 30 frames per second). Multiplying all these parameters together and dividing by 3 for the attenuating strips implies an A/D conversion rate of 4×10^7 points per second. A fully parallel system with one light valve and detector array for each bit plane could achieve 1.2×10^8 points per second.

An advantage of this technique is that it operates with incoherent input. The requirements on the spatial and temporal coherence of the readout illumination are sufficiently relaxed that noise problems associated with coherent spatial filtering or transforming techniques are avoided. The technique also minimizes the spatial-frequency requirements of the real-time device because the sharp edges of the binary dots in halftoning do not have to be maintained.

Although these initial results are encouraging, further application of the technique must await improved real-time devices. The aperiodic nature of the LCLV results in unequal quantization intervals and limits the number of bits. The LCLV device used in this experiment is inherently aperiodic because of the nonlinear response of the photoconductor and the nonlinear relationship between applied voltage and effective birefringence. Further developments in device technology may improve the periodicity and overall performance of the technique.

In a related application of a Hughes birefringent LCLV, Collins, Fatehi and Wasmundt [78] have implemented binary combinatorial logic operations by directly using the characteristic curves of the device. They have achieved two input logic functions such as AND, NOR, XOR, etc. by viewing logic as a nonlinearity similar to the procedure followed in the VGM logic described in Sect. 3.2. Their LCLV device was a parallel aligned, pure birefringent cell, similar to that used by Armand et al. [74] for A/D conversion. By appropriate choice of operating point on the device curves shown as dashed lines in Fig. 25 different functions are achieved. They use various rotations and combinations of polarizer/analyzer pairs to translate the curves left and right to match the desired characteristics. Their experimental results demonstrate several of the standard binary combinatorial logic functions.

Lee and coworkers have developed a system for obtaining logic operations using a liquid crystal optically addressed through a photoconductor matrix [33,34]. Each element of the array can perform simple combinatorial logic operations on two binary inputs.

As noted in Sect. 2.1, real time implementation of halftoning is limited by the lack of an optical input transducer with a threshold characteristic. The threshold itself is useful as a direct nonlinear function; for example the nonlinear curves needed for optical logic can be expressed as a sequence of thresholds. An attempt at improving the threshold characteristics of the Hughes LCLV has been described by Michaelson [79]. A feedback arrangement shown in Fig. 30 was used. As shown in the figure, a portion of the output light is directed back to the input of the light valve via a combination of beamsplitters and lenses. The light is summed with the input illumination at the surface of the light valve to produce positive feedback. Both the read and the write illumination were derived from an argon ion laser operating at 514.5 nm. It should be noted that although coherent light was used in the experimental procedure, the feedback arrangement and characteristics of the light valve are such that incoherent illumination would have worked as well. To avoid unwanted interference between the coherent input light and the feedback light, the input illumination was configured with its polarization orthogonal to the feedback component. In this manner, the incident light on the light valve was simply the sum of the intensities of the two components. In the initial state, with no input illumination and the feedback component set to zero by momentarily blocking the feedback path, the device remains in the off state resulting in zero output intensity. As the input illumination is increased, the device remains off until the threshold level of the light valve is reached. At this point, the device begins to turn on and as a result of the optical gain characteristic of the light valve and the positive feedback, regeneration occurs. The device is switched on with no further increase in input illumination. The regenerative process continues until a point is reached on the light valve transfer curve where the loop gain (device gain less the feedback losses) drops to unity. In practice this point is reached before the saturation level of the light valve is reached. As a result, the desired binary transfer function is not entirely achieved but rather a soft shoulder is obtained with further increase in input illumination. Experimental results are shown in Fig. 31 together with the light valve transfer function for comparison. As seen in the figure, the feedback system provides a very sharp threshold characteristic as well as a marked improvement in overall gamma.

The light valve used to obtain the data in Fig. 31 did not exhibit a particularly high gamma. Thus the output range over which regeneration occurred was somewhat limited. Further, device nonuniformities over the aperture precluded the possibility of testing the feedback arrangement with two-dimensional input images. To be effective, the device would

have to be sufficiently uniform to maintain the input threshold point within a reasonably narrow band over the area of the input image. Improved devices exhibiting both the necessary uniformity and increased initial gamma could render the feedback system a viable means of obtaining real-time sharp thresholding operation from the light valves.

In a related system, Sengupta, Gerlach and Collins [80] have used feedback around a LCLV to implement an optical flip-flop. Pairs of image elements on the LCLV are coupled by optical feedback so that latching at one of two stable states occurs. A parallel array of flip-flops is achieved with one LCLV by arranging the output of half the active image area to control the opposite half, and vice-versa. They present stability theory to predict the stable states and show experimental confirmation.

Another method of direct nonlinear processing uses several optically and electrically controlled liquid crystal light valves. This system has been described in detail by Michaelson and Sawchuk [81] and is called the multiple light valve system (MLVS). The system performs nonlinear processing in real-time and has the advantages of inconerent operation and electronic programming of arbitrary nonlinearities.

A functional schematic of the MLVS is shown in Fig. 32. The first element in the system maps the intensity variations of the two-dimensional input image into two-dimensional outputs with constant magnitude and temporal separation. This step is done by a time-scanning level slice produced with a LCLV. The constant magnitude outputs are then weighted as a function of time by another LCLV to give the desired nonlinear transformation. The weighted outputs are then integrated over an appropriate time interval to give a nonlinearity transformed output image. The operation is shown schematically in Fig. 32 for a simple three intensity level, two-dimensional input image. The intensity-to-time converter maps input intensities over the range of 0 to I_{max} into the time interval [0,T]. Thus for a linear mapping, the output planes for the intensity-to-time converter will remain dark during the time interval $0<t<t_1$ since there are no intensity components in the input image between $I = 0$ and $I = I_1$. At time t corresponding to the input intensity I_1, the output plane will have a constant intensity response over all portions of the input image for which $I = I_1$. Similar output responses occur at times t_2 and t_3 as shown in the figure. For all other times $0<t<T$ with $t \neq t_1,t_2,t_3$ the output intensity remains at zero. The temporal intensity weighter is simply an electrooptic attenuator which weights the constant intensity, time sequential outputs of the intensity-to-time converter in the desired nonlinear manner. The time sequential weighted responses are then summed over the interval [0,T] in the integrator. Thus at time $t = T$, a nonlinearly transformed input image will be present at the integrator output. A simple logarithmic compression is depicted in Fig. 35. In practice, the integrator can be a television system, photographic film, or another real-time image transducer.

One of the key features of the system is the ability to arbitrarily change the form of the nonlinear function in real time. The nonlinearity is introduced by applying a nonlinearly shaped voltage waveform into the temporal intensity weighter during the time interval [0,T]. In the system being evaluated, the waveform is produced by a microprocessor-D/A converter and can arbitrarily be changed to effect numerous transformations including logarithm, exponentiation, level slice, and A/D conversion. Details of implementation and experimental measurements are given in Ref. [79,81].

Figures 33 and 34 show operation of the system with a logarithmic and exponential characteristic. Figure 35(b) shows the experimental results of a level slice operation performed on the variable intensity input image of Fig. 35(a).

Farhat [35] has discussed the possibility of using a photochromic material to implement a direct nonlinearity. In particular, he suggests the possibility of performing division optically by placing a transparency in contact with a photochromic material having a negative response characteristic. In his paper, Farnat also suggests several variations of composite nonlinear systems obtained by placing a photochromic nonlinear element at various places in an optical processing system

A final technique for direct point nonlinearities has been developed by Santamaria et al. [32] for providing a square root nonlinearity. The method does not fit into any of the three rubrics listed above under point nonlinearities. The method consists of doing a time-sequential reconstruction of an image by imaging through a filter that transmits the zero order and one other spatial frequency component. The selection of the second component is done by scanning an aperture in the Fourier plane.

5. Prospects for future research

Finally, a few words about future directions for nonlinear processing and optical processing in general can be given. One of the major limitations of optical processing in

comparison to digital systems is the limited accuracy available. Closely related issues are the inherent system noise and dynamic range; in fact, accuracy can be thought of as the number of distinguishable signal levels within the usable dynamic range. For analog optical processing, 6 bit accuracy (64 levels) is moderately difficult, 8 bit accuracy (128 levels) is very difficult, and 10 bit accuracy (1024 levels) may be impossible. For the human observer, 6 bits may be acceptable for the final output of an image processing system, but more accuracy may be needed internally. Digital computers can have internal processing registers of 64 bits or more, so that their accuracy is eventually limited by the accuracy of the sensors which convert the analog input to digital form. Accuracy, noise and dynamic range problems become more serious with nonlinear processing. Signal-independent noise becomes signal-dependent, more internal processing accuracy is needed to accommodate larger variations in signal magnitude, scaling of quantities internally is hard to predict, and analysis of these effects becomes difficult. Experimental and theoretical research to study these effects is needed, particularly in the context of optical processors.

Research on optical real-time devices with novel operations and threshold behavior should be supported in the future. Devices such as the variable-grating-mode (VGM) devices [70,71] that perform intensity-to-spatial frequency conversion can perform analog nonlinear processing in a flexible manner and also can perform binary or residue [22-26] arithmetic. Research on VGM devices is in the early stages and very little thought has been given to the application of this new type of element in optical processors.

Related to this is current research on numerical optical processors that use different number systems (binary or residue) with a relatively small number of well-defined distinguishable levels. For binary systems, the base is, of course, 2 levels; for residue systems the base may be on the order of 20. To perform this type of optical processing in real time, new optical devices that exhibit threshold behavior must be developed. In this paper, the work on optical logic and feedback for threshold sharpening are preliminary efforts in this direction. Recent work on microchannel plate optical devices [83,84] and cellular liquid crystal systems [85] also show promise. In effect, real-time devices with parallel optical bistability, fast switching times (on the order of milliseconds or less), good resolution (a 1000x1000 point array is desirable) are needed. The advantages of signal regeneration, noise suppression and logical operations are well known in serial digital processors; the hardware and software implications of a parallel optical logic architecture remain almost completely unknown and should be investigated in the future.

6. Summary and conclusions

This paper has attempted to review briefly work in nonlinear optical information processing. The field is broad, and in the limited space available we have concentrated on point nonlinear functions because they are the best developed theoretically and experimentally. The main approaches to point nonlinear processing have been: halftoning; intensity-to-spatial frequency conversion; and direct nonlinear processing using the inherent characteristics of image detection and modulation systems. Real-time implementations of all these approaches along with many application examples have been discussed. Other areas which have not been covered in detail will be included in a greatly expanded version of this paper to be published [29]. Additional topics to be included are: a classification of nonlinear systems based on the cascade of linear systems with point nonlinear point functions; feedback optical processors which perform nonlinear functions; details of halftone screen synthesis procedures; analysis of the degrading effects of non-ideal halftone recording media; and design examples for many specific types of useful nonlinearities.

Acknowledgements

The authors wish to acknowledge the support of the Air Force Office of Scientific Research, Electronics and Solid State Sciences Division, under Grant AFOSR-77-3285, the National Science Foundation, under Grant ENG76-15318 and ENG78-11368, and the Joint Services Electronics Program (AFSC) at USC.

The authors are grateful to A. Armand, P. Chavel, S. Dashiell, J. Michaelson, B.H. Soffer and A.R. Tanguay, Jr. for their contributions to the research described in this paper.

References

1. Andrews, H.C., Tescher, A.G. and Kruger, R.P. (1972). IEEE Spectrum. 9, No. 7, 20-32.

2. Hunt, B.R. (1975). Proc. IEEE. 63, 693-708.

3. Casasent, D., and Sterling, W. (1975). IEEE Trans. Computers. C-24, 348-358.

4. Casasent, D., and Psaltis, D. (1980). Opt. Lett. 5, 395-397.

5. Hester, C.F., and Casasent, D. (1980). Appl. Opt. 19, 1758-1761.

6. Goodman, J.W. (1968). "Introduction to Fourier Optics." McGraw-Hill, New York.

7. Lee, S.H. (1974). Opt. Eng. 13, 196-207.

8. Goodman, J.W. (1977). Proc. IEEE. 65, 29-38.

9. Casasent, D. (1977). Proc. IEEE. 65, 143-157.

10. Landauer, R. (1976). In "Optical Information Processing" (Y.E. Nesterikhin et. al., eds.), pp. 219-253. Plenum Press, New York.

11. Taylor, H.F. (1978) Appl. Opt. 17, 1493-1498.

12. Goldberg, L., and Lee, S.H. (1979). Appl. Opt. 18, 2045-2051.

13. Schaefer, D.H., and Strong, J.P., III. (1977). Proc. IEEE. 65, 129-138.

14. Athale, R.A. and Lee, S.H. (1979). Opt. Eng. 18, 513-517.

15. Preston, K., Jr. (1972). "Coherent Optical Computers," ch. 8, McGraw-Hill, New York.

16. Guest, C.C., and Gaylord, T.K. (1979). In "SPIE Proc. on Optical Processing Systems," Vol. 185, pp. 42-49.

17. Cederquist, J., and Lee, S.H. (1979). Appl. Phys. 18, 311-319.

18. Akins, R.P., Athale, R.A., and Lee, S.H. (1980). Opt. Eng. 19, 347-358.

19. Collins, S.A., Jr., Gerlach, V.H., and Zakman, Z.M. (1979). In "SPIE Proc. on Optical Processing Systems," Vol. 185, pp. 36-41.

20. Garmire, E., Marburger, J.H., and Allen, S.D. (1978). Appl. Phys. Lett. 32, 320.

21. Kruse, B. (1978). AFIPS Conf. Proc, 47, 1015-1024.

22. Huang, A. (1973). IEEE Trans. Computers C-22, 14-18.

23. Huang, A., Tsunoda, Y., Goodman, J.W., and Ishihara, S. (1979). Appl. Opt. 18, 149-162.

24. Psaltis, D., and Casasent, D. (1979). Appl. Opt. 18, 163-171.

25. Horrigan, F.A., and Stoner, W.W. (1979). In "SPIE Proc. on Optical Processing Systems," Vol. 185, pp. 19-27.

26. Tai, A., Cindrich, I., Fienup, J.R., and Aleksoff, C.C. (1979). Appl. Opt. 18, 2812-2823.

27. Collins, S.A., Jr. (1977). In "SPIE Proc. on Effective Utilization of Optics in Radar Systems," Vol. 128, pp. 313-319.

28. Collins, S.A., Jr., Ambuel, J., and Damon, E.K. (1978). In "Proc. ICO-II Conf.," pp. 311-314. ICO, Madrid.

29. Sawchuk, A.A., and Strand, T.C. (to be published). In "Applications of Optical Fourier Transforms." (H. Stark, ed.), Academic Press, New York, Chp. 9.

30. Cederquist, J., and Lee, S.H. (1975). Appl. Phys. 18, 311-319.

31. Bartholomew, B.J., and Lee, S.H. (1980). Appl. Opt. 19, 201-206.

32. Lee, S.H. (1978). In "Optical Information Processing." Y.E. Nesterikhin, G.W. Stroke and W.E. Kock (eds.). Plenum Press, New York. 255-279.

33. Lee, S.H. (1978). In "Optical Information Processing," Vol. 2, E.S. Barrekette, G.W. Stroke, V.E. Nesterikhin, and W.E. Kock. (eds.). Plenum Press, New York, 171-191.

34. Lee, S.H. (to be published). In "Optical Information Processing Fundamentals." S.H. Lee (ed.) Springer-Verlag. Chp. 7.

35. Farhat, N.H. (1975). IEEE Trans. on Computers. C-24. 443-448.

36. Kato, H., and Goodman, J.W. (1973). Opt. Communications. 8, 378-381.

37. Kato, H., and Goodman, J.W. (1975). Appl. Opt. 14, 1813-1824.

38. Strand, T.C. (1975). Opt. Communications. 15, 60-65.

39. Dashiell, S.R., and Sawchuk, A.A. (1975). Opt. Communications. 15, 66-70.

40. Dashiell, S.R., and Sawchuk, A.A. (1977). Appl. Opt. 16, 1009-1025.

41. Armand, A. (1979). "Real-Time Nonlinear Optical Information Processing," Ph.D. Thesis, Department of Electrical Engineering, University of Southern California, Los Angeles, Ca. 90007, USCIPI Report 880.

42. Dashiell, S.R., and Sawchuk, A.A. (1977). Appl. Opt. 16, 1936-1943.

43. Matsumoto, S., and Liu, B. (1979). Appl. Opt. 18, 2792-2802.

44. Beard, T.D., Bleha, W.P., and Wong, S-Y. (1974). Appl. Phys. Lett. 22, 90.

45. Grinberg, J., Jacobson, A., Bleha, W., Miller, L., Fraas, L., Boswell, D., Myer, D. (1975). Opt. Eng. 14, 217-225.

46. Iwasa, S., and Feinleib, J. (1974), Opt. Eng. 13, 225.

47. Bleha, W., Lipton, L., Wiener-Avnear, E., Grinberg, J., Reif, P., Casasent, D., Brown, H.B., and Markevitch, B. (1978). Opt. Eng. 17, 371-384.

48. Dashiell, S.R., and Sawchuk, A.A. (1977). Appl. Opt. 16, 2279-2287.

49. Abbe, E. (1893). Archiv. Mikroskopische Ant. 9, 413.

50. Marquet, M. (1959). Opt. Acta 6, 404-405.

51. Marquet, M., and Tsujiuchi, J. (1961). Opt. Acta 8, 267-277.

52. Delingat, E. (1973). Optik 37, 82-90.

53. Pappu, S.V., Kumar, C.A., and Mehta, S.D. (1978). Curr. Sci. 47, No. 1, 1-6.

54. Roychoudhuri, C., and Malacara, D. (1975). Appl. Opt. 14, 1683-1689.

55. Sawchuk, A.A., and Dashiell, S.R. (1975). In "Proc. IEEE 1975 International Optical Computing Conference", Washington, D.C., pp. 73-76.

56. Schwider, J., and Burow, R. (1970). J. Opt. Soc. Amer. 60, 1421.

57. Delingat, E. (1972). Optik 34, 433-441.

58. Schneider, W. (1974). Optica Acta. 21, 563-576.

59. Schnieder, W., Fink, F., and Van Der Piepen, H. (1975). Optcs Commun. 14, 42-45.

60. Liu, H-K., Goodman, J.W., and Chan, J. (1976). Appl. Opt. 15, 2394-2399.

61. Lohmann, A., and Strand, T.C. (1975). In "Proc. Electro-Optical Systems Design - 1975." Anaheim, California. pp. 16-21.

62. Liu, H-K. (1978). Appl. Opt. 17, 2181-2185.

63. Strand, T.C. (1976). "A Method of Nonlinear Optical Information Processing," Ph.D. Thesis, Department of Applied Physics and Information Science, University of California-San Diego, La Jolla, Ca. 92093.

64. Liu, H.K. (1978). Opt. Lett. 3, 244-246.

65. Sawchuk, A.A., and Dashiell, S.R. (1976). In "SPIE Proc. Optical Information Processing," Vol. 83, pp. 130-136.

66. Liu, H-K., and Goodman, J.W. (1976). Nouv. Rev. Optique. 7, 285-289.

67. Indebetouw, G. (1977). Appl. Opt. 16, 1951-1954.

68. Morgenstern, B., and Lohmann, A.W. (1963). Optik 20, 450-455.

69. Armitage, J.D., and Lohmann, A.W. (1965). Appl. Opt. 4, 399-403.

70. Armand, A., Sawchuk, A.A., Strand, T.C., Boswell, D., and Soffer, B.H. (1978). In "Proc. 1978 Int'l Optical Computing Conference." pp. 153-158. London.

71. Armand, A., Sawchuk, A.A., Strand, T.C., Boswell, D., and Soffer, B.H. (1978). In "Proc. International Commission for Optics - 11." pp. 253-255. Madrid.

72. Soffer, B.H., Boswell, D., Lackner, A.M., Tanguay, A.R., Jr., Strand, T.C., and Sawchuk, A.A. (1980). In "SPIE Proc. Devices and Systems for Optical Signal Processing," Vol. 218, pp. 81-87. Los Angeles.

73. Soffer, B.H., Boswell, D., Lackner, A.M., Chavel, P., Sawchuk, A.A., Strand, T.C., Tanguay, A.R., Jr. (1980). In "Proc. 1980 Int'l Optical Computing Conf.", vol. 232, Book II, pp. 128-136. Washington, D.C.

74. Armand, A., Sawchuk, A.A., Strand, T.C., Boswell, D., and Soffer, B.H. (1980). Opt. Letters 5, 129-131.

75. Tai, A., Cheng, T., and Yu, F.T.S. (1977). Appl. Opt. 16, 2559-2564.

76. Agfa-Gevaert Inc., 275 North Street, Teterboro, New Jersey 07608 USA.

77. Grinberg, J. and Marom, E. (1977). In "SPIE Proceedings on Optical Signal and Image Processing," Vol. 118, pp. 66-74.

78. Collins, S.A., Jr., Fatehi, M.T., and Wasmundt, K.C. (1980). In "SPIE Proc. 1980 International Optical Computing Conf." Vol. 232, Book 2, pp. 168-173.

79. Michaelson, J. (1979). "Characterization of Liquid Crystal Light Valves and Their Applications to Real-Time Nonlinear Optical Processing," Ph.D. Thesis, Department of Electrical Engineering, University of Southern California, Los Angeles, Ca. 90007, USCIPI Report 930.

80. Sengupta, U.K., Gerlach, U.H., and Collins, S.A., Jr. (1978). Opt. Lett. 3, 199-201.

81. Michaelson, J., and Sawchuk, A.A. (1980). In "SPIE Proc. Devices and Systems for Optical Signal Processing," Vol. 218, pp. 107-115, Los Angeles.

82. Santamaria, J., Ojeda-Castaneda, J., and Berriel-Valdes, L.R. (1980). In "Proc. International Commission for Optics," Ensenada, Mexico.

83. Warde, C., Fisher, A.D., Cocco, D.M., and Burmawi, M.Y. (1978). Opt. Lett. 3, 196-198.

84. Warde, C., Weiss, A.M., Fisher, A.D., and Thackara, J.I. submitted to Appl. Opt.

85. Athale, R.A., and Lee, S.H. submitted to Appl. Opt.

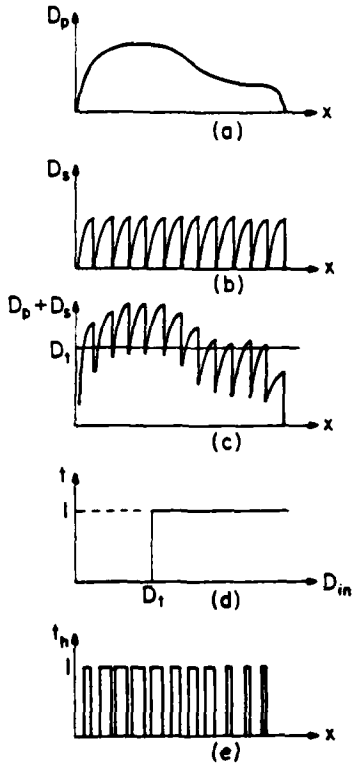


Figure 1. The halftone encoding process. A continuous input density (a) adds with the halftone screen density (b) to produce a modulated image density (c). This is recorded with a hardclipping film whose characteristic is shown in (d). (e) is the resultant halftoned image.

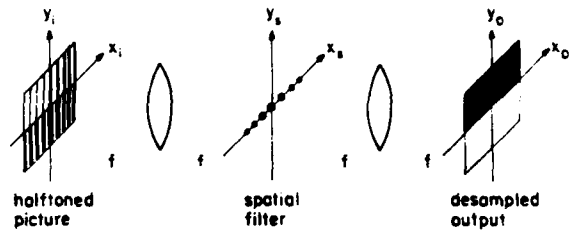


Figure 2. Demodulation of the halftoned picture in a coherent optical system.

Figure 3. Transmittance vs. input density for a general non-ideal recording

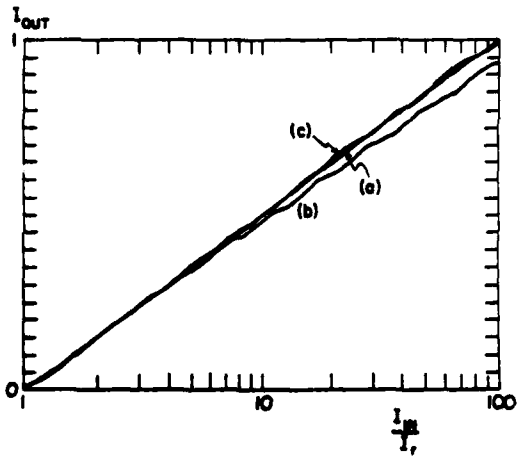


Figure 4. Logarithmic function for a piecewise linear model of a recording medium with gamma = 3.0. (a) Ideal. (b) Degraded. (c) Optimized.

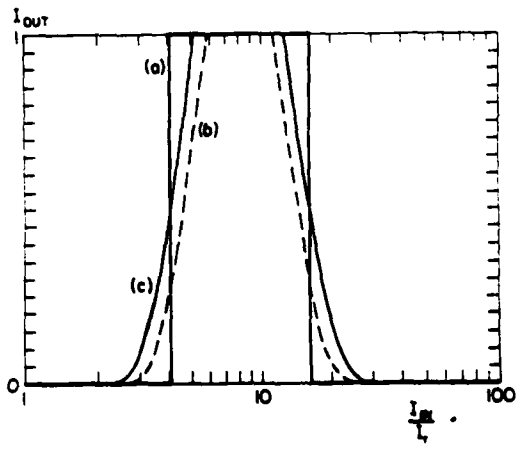


Figure 5. Level slice function for a piecewise linear model of a recording medium with gamma = 3.0. (a) Ideal. (b) Degraded. (c) Optimized.

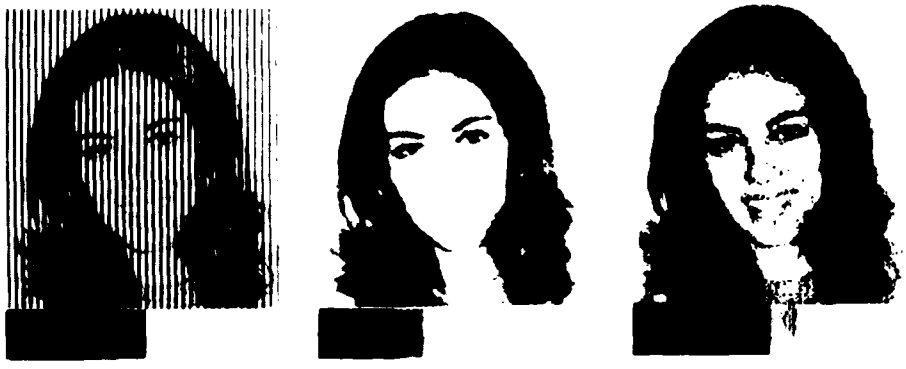


Figure 6. Logarithmic filtering with photographic film recording. (a) The original input pattern is a girl's face multiplied with a grating. The dynamic range is from 0 to 2 in density. (b) Logarithmic filtering to remove the grating. (c) Linear filtering to remove the grating. (From Kato and Goodman [36,37]).

Figure 7. Level slice function with photographic film recording. (a) Continuous tone input. (b) Level slice output.

Figure 8. Isophote generation with photographic film recording. (a) Input image. (b) Isophote output. Each isophote represents a doubling of input amplitude.

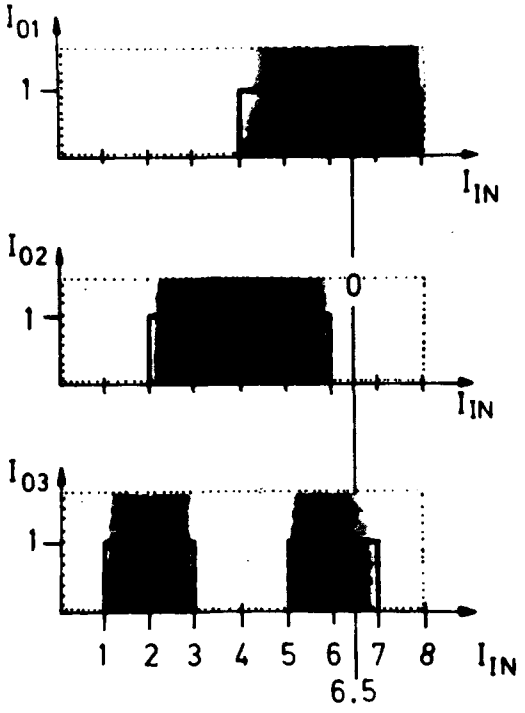


Figure 9. First three bit-planes of a Gray code A/D conversion performed on a linear intensity wedge.

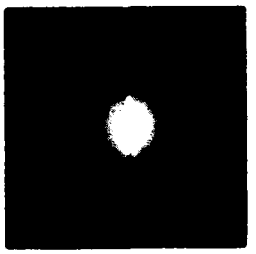


Figure 10. Input image for optical A/D conversion.

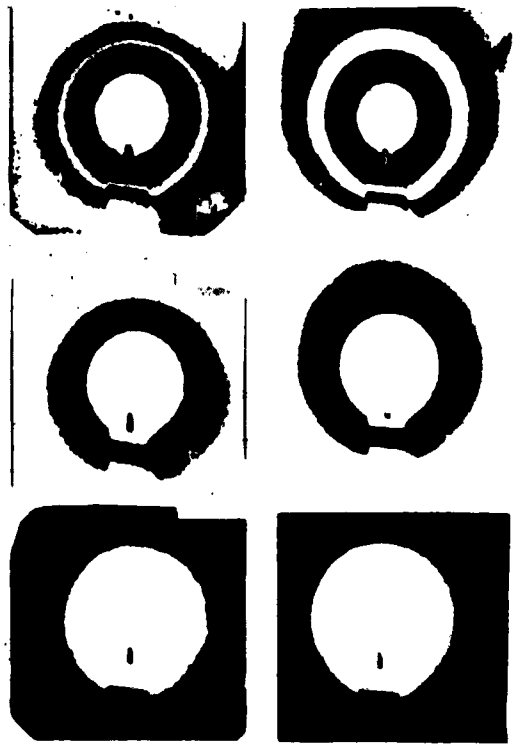


Figure 11. First three bit planes of a Gray code A/D conversion. The bit planes on the left-hand side were generated optically. The bit planes on the right-hand side were produced digitally after scanning with a microdensitometer.

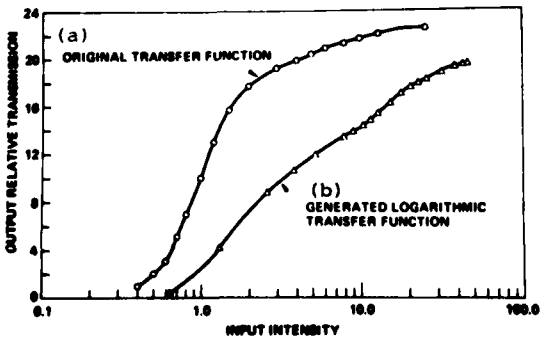


Figure 12. Liquid crystal characteristic curve. (a) Without the halftone screen. (b) With halftone screen designed to give a logarithmic response over two decades.

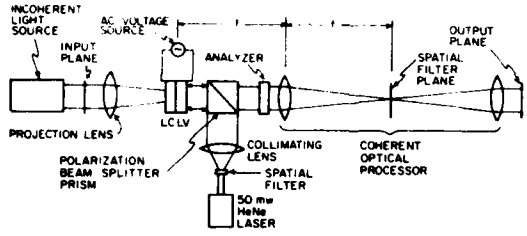


Figure 13. Experimental setup for real-time halftone processing with a LCLV.

(a)

(b)

Figure 15. Real-time homomorphic filtering. (a) Input image is a face with simulated push-broom scanner noise. Overall density range is 2.0 D. (b) Homomorphic filtered output.

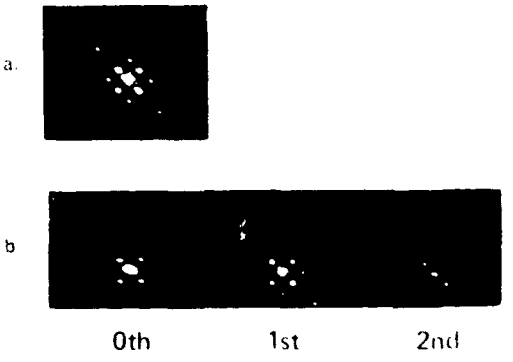


Figure 14. Real-time logarithmic processing with a halftone screen and a LCLV. (a) Fourier transform of two crossed gratings imaged on the LCLV with no halftone screen. The grating spectra are convolved with one another. (b) Fourier transform of crossed gratings with the halftone screen. In the zeroth order, the off-axis terms have been eliminated indicating the grating spectra are not convolved with one another.

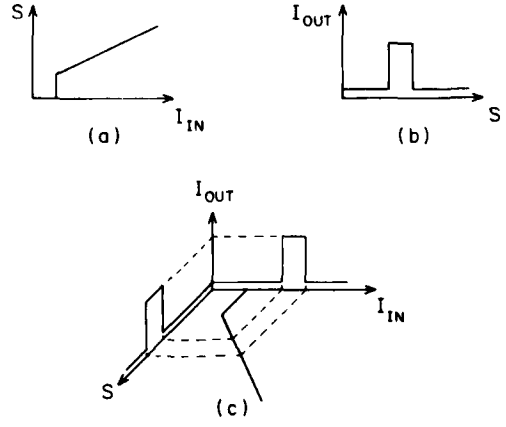


Figure 16. Intensity to spatial frequency conversion for nonlinear processing. (a) A device is used to convert input intensity into a varying spatial frequency s. (b) A Fourier plane filter selectively attenuates the spatial frequency component s. (c) The I_{out} vs. I_{in} characteristic of the overall filtering system can be found graphically by tracing through the two characteristics of (a) and (b).

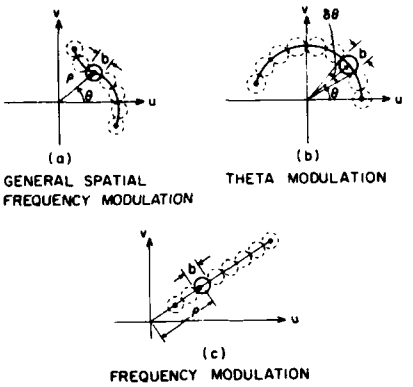


Figure 17. Intensity-to-spatial frequency conversion (a) General case. (b) Theta modulation. (c) Frequency modulation.

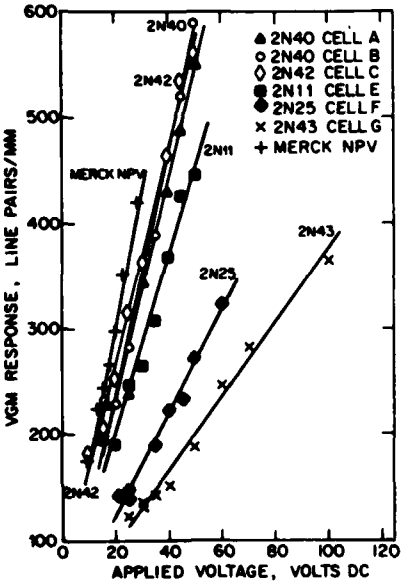


Figure 19. VGM spatial frequency as a function of applied voltage.

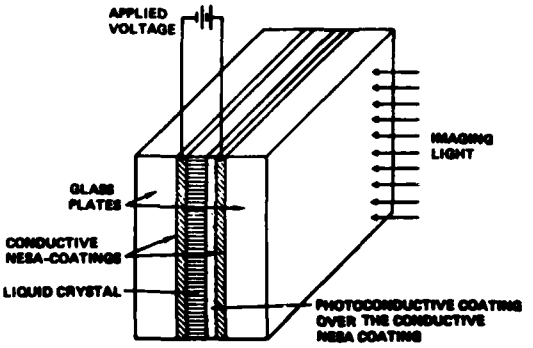


Figure 18. Construction of an optically activated VGM real-time device.

Figure 20. The phase grating structure of the VGM device viewed through a phase contrast microscope.

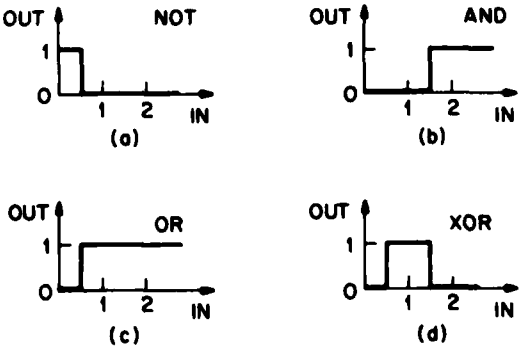


Figure 21. Logic functions as binary nonlinearities. Given an input consisting of the sum of two binary inputs, different logical operations can be effected on those inputs by means of the depicted nonlinearities. (a) NOT, (b) AND, (c) OR, (d) XOR.

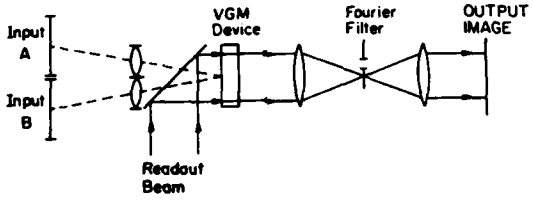


Figure 22. Experimental arrangement for performing logical operations on two-dimensional, binary inputs with a VGM device. The two binary input images are superimposed on the photoconductor. The device is read out in transmission. Simple slit filters can be used to achieve the desired logic operations.

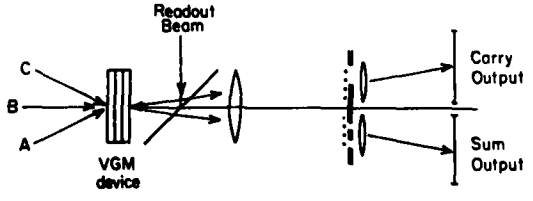


Figure 24. Implementing sum and carry bit planes of a full adder with one VGM device. The inputs are bit-planes from images A and B and the carry bit-plane C. Whenever two or more of the three inputs are "1", the carry output should be "1". The sum bit should be "1" if an odd number of inputs are "1".

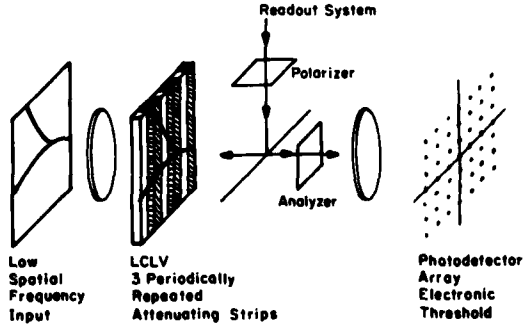


Figure 26. System for parallel A/D conversion.

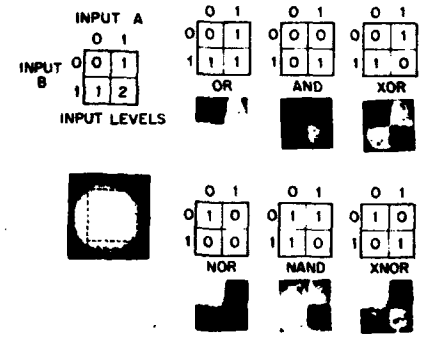


Figure 23. VGM logic results. Two binary images were superimposed on the photoconductor to produce the input intensities as shown in the upper-left corner. Without any filter, the output is ideally a uniform field (logical 1). The output field with no filter is shown in the lower left corner with the image area of interest indicated by the dotted lines. The truth tables for the various logic functions are shown on the right above their corresponding experimental output.

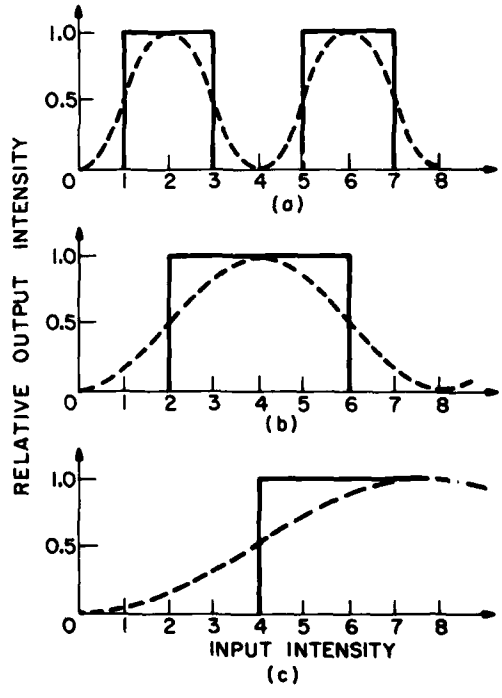


Figure 25. Nonlinear characteristic required for the three-bit Gray code. Solid curves are the desired characteristics for the bit-plane outputs. Dotted curves are the ideal sinusoidal response of a linear birefringent device. Parts (a) through (c) represent increasingly significant output bits.

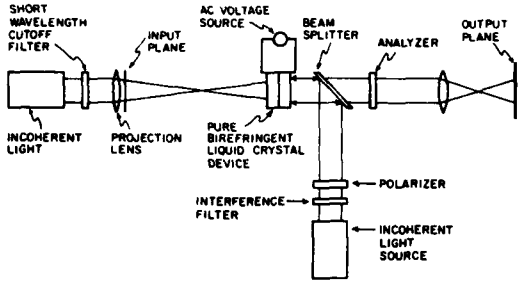


Figure 27. Experimental setup for real-time, parallel A/D conversion.

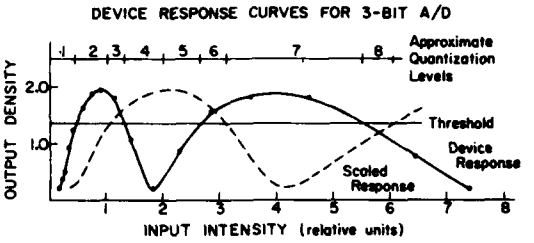


Figure 28. Response curve of the liquid-crystal device used for the three-bit A/D conversion. The solid curve is the measured response. The dotted curve represents the same response with a fixed attenuation of the input.

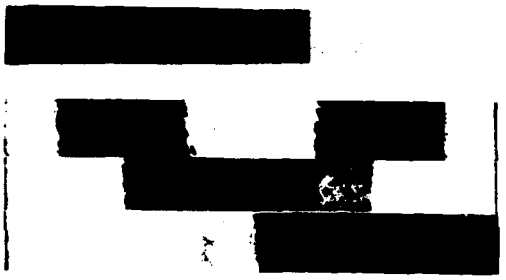


Figure 29. Direct A/D conversion. The eight-level analog input is shown at the top. Below is the binary coded output in the form of three bit planes of the Gray code.

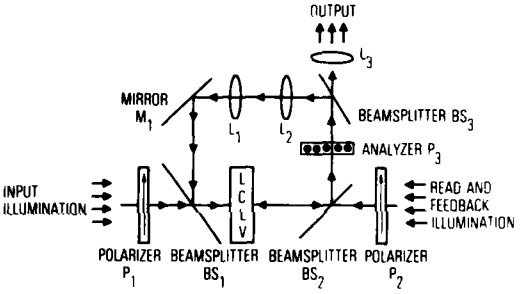


Figure 30. Experimental optical feedback system using a LCLV.

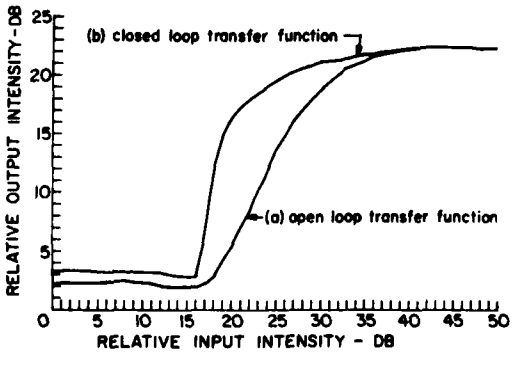


Figure 31. LCLV feedback system response.

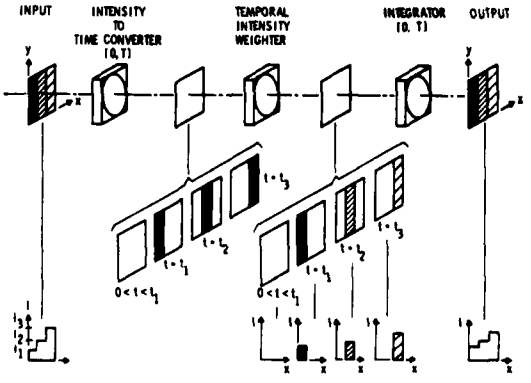


Figure 32. Schematic diagram of the multiple light valve system (MLVS) for nonlinear optical processing.

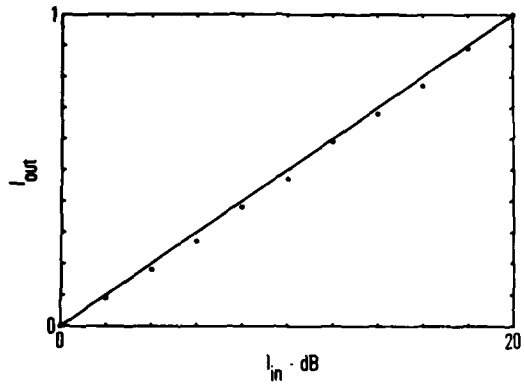


Figure 33. Response of the MLVS programmed for a logarithmic characteristic.

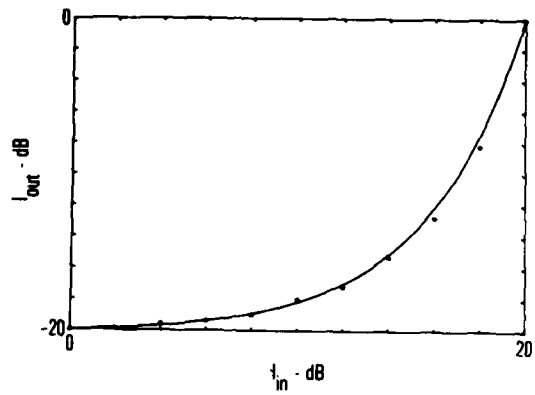


Figure 34. Response of the MLVS programmed for an exponential characteristic.

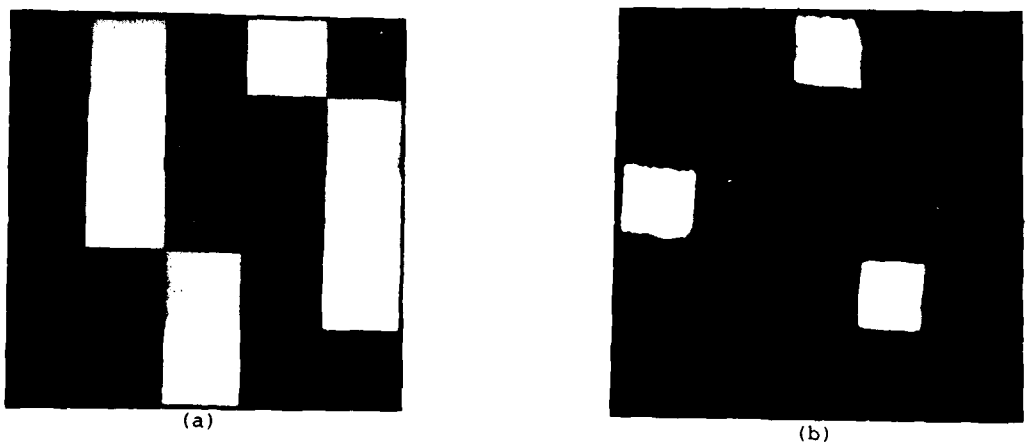


Figure 35. Level slice operation performed with the MLVS. (a) Input image. (b) Level slice output.

Discussion (Alexander A. Sawchuk; Discussion Leader: Hua-Kuang Liu)

Q. You've given a very nice summary of point nonlinearitites and their role. Could you make a comment about non-point processing, such as perhaps median filtering? Is there a future for that?

A. The answer is that if you can do logic, then you can do median filtering, so it doesn't really matter if you are working on a point or a total neighborhood of points. If you can make decisions (e.g. is this sample greater than that sample), then you can effectively do a median filter.

R. We can do that with parallel processors by using feedback.

C. There is a meeting going on now in North Carolina on optical bistability. It is already possible to make a bistable device with spatial resolution. Now it is only 4 or 6 elements, but the approach does exist. The whole area of nonlinear optics is directed towards developing a nonlinear spatially resolved device.

R. That's good, because if we can get an imaging optical device with a reasonable number of resolution elements, even 100 x 100, then we can do a lot of these things and do them adequately.

C. If you can do more than just a point with optics, then you have an advantage over electronic chips in the connectivity area. They can easily do one pixel at a time, but if you want to look at several neighbors, you have to have wires to do that connection. Thus the neighborhood in VLSI has to be rather small. That is one problem we might avoid in optics because the interconnections are easier.

C. I'm curious about the variable grating mode used in the light valve. It seems you mentioned that the spatial frequency bandwidth extended from about 100 lines/mm to 800 lines/mm. That's more than you can exploit in one dimension.

R. It does lead to problems because your optics must be quite good to resolve these high orders. You have to be careful about the resolution number because although the grating period is, say, 600 lines/mm, each pixel must extend over several grating periods, so we are really talking about 5 or 6 periods within each pixel.

C. One important point about doing analog processing in the variable grating domain: if you want to get on the order of 10 quantization levels, you need to have a number of grating periods per pixel of 10 or more.

C. The A/D converter scheme that we have shown in this talk is done with bi-refringent valves, it is not done with the variable grating mode. The fundamental limitation there in terms of the number of bits we can obtain is one of the periodicity of the valve itself. The valve is really two components, there's the bi-refringent component and a photoconductive part. We believe that most of the nonlinear effects are due to the photoconductor, so we really feel it is a photoconductor problem.

C. I was thinking of the variable grating mode and how you are using it, in terms of how one might come up with a new invention. The critical thing you have done is perform a mapping from intensity to position in a plane, in this case a spatial frequency plane, which you can access with filters. That particular mapping involves intensity to spatial frequency and then to spatial position. Maybe that suggests looking at other intermediate steps where you go, say, from intensity to color and from color to spatial position.

Q. Explain what happens mathematically with the variable grating system.

A. Every resolvable input brightness level locally modulates the variable grating. You then get different orders in the spatial frequency plane, and the positional information of each pixel is modulated on top of its respective order. When you inverse transform, the pixel comes out in its correct spatial orientation but with a change in gain which you can control by putting an attenuating filter in the spatial frequency plane. This is a new kind of device and there are many new kinds of things you can do with it. There may be many applications to space-variant systems because this is an intensity variant processing scheme.

Q. You mentioned that 6 to 8 bits of accuracy is what we seem to be able to do with analog systems, and that seems to agree with everybody's experience. In the field of analog electronic computers, you find much higher accuracies possible, down in the 1 part in 10,000 range. What are your views on why we seem to have an accuracy floor that is so different from what is encountered in analog electronic processing?

A. I think one problem is that in optics you have a parallel array of channels, and in an analog computer you essentially have one dimension. Another thing is that the inherent nature of electronics is much more noise free than coherent optics, where we have to worry about phase and other things. Also, if you look at techniques other than intensity, like polarization, for carrying information, it is rather difficult to measure polarization to 1 part in even 20.

C. I have a speculation that these electronic analog computers usually have feedback to stabilize the system, and none of our optical systems have feedback.

Q. Do we really need that kind of accuracy in our systems?

A. Yes, for people who are not so much interested in looking at pictures at the output, but rather discriminating a complicated signal with large dynamic ranges.

C. Electronic devices are also inherently much more linear than most of the things we talk about in optics. Almost every time you do an image intensity to electronic signal transfer, you are talking about working with very nonlinear devices, and there is no single device you can think of that has a very wide range of linearity except a photomultiplier.

C. Also PIN junctions as detectors have large dynamic ranges and good linearity, even arrays of them.

C. There's no physics in talking about parallel versus single channels and saying you ought to have more noise in parallel systems.

R. Yes, there is, because of crosstalk problems.

. But isn't that a scattering problem?

A. No. When you have a PIN photodiode, the reason that it is linear is that the device is fully depleted. If you do the thing in parallel, you have complete crosstalk between the channels, there is no way to isolate unless one would array them. In trying to array them, one gets into a very difficult processor problem. So I think there is some difference there between parallel and single channel processors.

C. It's a very serious problem if the accuracy of an optical system is limited to 6 to 8 bits. Then you can forget about doing most radar signal processing, because they want 90 db of dynamic range. There are of course lots of other applications without this constraint. But I would like to know the basic physics behind this limit. If there is no physics, then there is research to be done. If there is a physics limitation, then people don't need to spin their wheels on this type of application.

C. I'd like to mention that with a good liquid crystal light valve and an image converter to increase gain, you can use a feedback system which would give you the linearity that you need.

SPATIAL LIGHT MODULATORS FOR REAL TIME OPTICAL PROCESSING

Armand R. Tanguay,Jr.
Image Processing Institute
Departments of Electrical Engineering
and Materials Science
University of Southern California
University Park, Los Angeles, California 90007

Abstract

Several candidate real time spatial light modulator technologies for coherent optical processing applications are reviewed. Physical principles of operation are described, as are current technological and fundamental physical limitations on device performance. A number of promising directions for current and future research on spatial light modulators are presented.

I. Introduction

Following the advent of the laser, and directly stimulated by its ready availability, tremendous progress has been achieved during the past two decades in the development of coherent and incoherent optical processing techniques. However, the many advantages of two-dimensional parallel processing are not fully exploitable in image and data processing systems without the availability of appropriate real time spatial light modulators. Although a wide range of candidate spatial light modulator technologies have been proposed and extensively researched and developed, no single candidate has as of yet emerged that at once satisfies the system requirements of low cost, reliability, and ease of operation while simultaneously exhibiting the requisite technical performance characteristics demanded by increasingly complex optical processing applications.

Several factors have primarily contributed to the present gap between the level of sophistication of optical processing techniques and the development status of spatial light modulators necessary for real time implementations of such techniques. First and foremost, device research and development is inherently expensive due to large capital equipment requirements, necessary parallel research and development efforts in growth, deposition, and characterization of materials, and multiple iterations of device design on the basis of operational characterization and application-dependent criteria. To date, the prospective market for such devices has not generally been viewed as large enough to encourage the type of major investments that stimulated rapid progress in such areas as information displays and solid state lasers. An added complication arises from the fact that all of the current candidate spatial light modulators are active devices, and as such require extensive parametric characterization coupled with comprehensive understanding of the fundamental device physics for optimum results in diverse applications. Major progress in both of these areas has been a relatively recent development. Finally, it has become increasingly apparent that no single spatial light modulator (at least of those presently envisioned) is capable of satisfying the wide range of at times conflicting demands stemming from the great diversity of proposed real time optical processing applications.

In the past few years, research has been initiated in a number of laboratories on the physical principles of operation of generic classes of spatial light modulators, on the requisite materials technologies, on methods of device characterization and analysis, and on new types of spatial light modulators for special applications. Such research has demonstrated encouraging results, particularly in the areas of increased physical understanding of optimum device operational modes and design parameters, and of novel device technologies.

In view of the numerous extensive reviews of spatial light modulators that have already been published (1-7) the purpose of this paper is to present a survey of recent results in the research areas described above. In particular, emphasis will be placed on our present understanding of some of the fundamental physical limitations inherent in several of the more promising spatial light modulator technologies. In succeeding sections of this paper, performance parameters of spatial light modulators are briefly described with emphasis on several difficulties inherent in such performance evaluation and specification. Current progress in three major types of spatial light modulators (electrooptic, liquid crystal, and photorefractive) is then described. Finally, some future directions for research in several of these areas are suggested.

II. Performance Evaluation of Spatial Light Modulators

A wide range of potential applications exists for real time, recyclable spatial light modulators in systems implementations of coherent optical processing techniques. These applications include use as the input incoherent-to-coherent image transducer in optical correlators and convolvers, programmable Fourier plane filters, serial-to-parallel (two-dimensional) buffer memories, page composers for holographic memories, real time holographic recording media, and holographic Fourier plane filters. In addition, many optical processing applications demand considerable image preprocessing, including contrast variation, contrast enhancement, contrast reversal, edge ennancement, image addition/subtraction, thresholding, level slicing, and minimization or elimination of the zeroth diffracted order in the Fourier plane.

Performance evaluation of each of the candidate spatial light modulator technologies is complicated by the innerently diverse requirements demanded by such a wide range of applications and desirable features. A number of performance parameters important in coherent optical processing applications are presented in Table I. This listing characterizes device parameters required for primarily "linear" applications, in which the SLM output amplitude is optimally a linear function of tne input intensity. Additional parameters should be added for characterization of SLM's designed for "nonlinear" applications, including ideality of the implemented nonlinear function (e.g., logarithmic for homomorphic filtering, or step for thresholding), programmability of the nonlinearity (e.g., variable level slicing), extended resolution requirements (e.g., halftoning) and the necessity for post-modulator thresholding (i.e., electronic or optical).

TABLE I: PERFORMANCE PARAMETERS FOR SPATIAL LIGHT MODULATORS: "LINEAR" APPLICATIONS.

linearity	cycle time
sensitivity	optical quality
resolution; MTF	optical uniformity
contrast	phase dependence
address mechanism	operational complexity
write/read wavelengths	fabrication complexity
erase mechanism	lifetime
grey scale (dynamic range)	cost
storage capability/time	special features
phase or amplitude readout	reciprocity behavior
Fourier plane signal-to-noise ratio	

The difficulty of side-by-side comparison of SLM's is readily apparent from the considerable number of characteristics presented in Table I. The suitability of an individual SLM also depends critically on the particular application for which it is envisioned. Furthermore, data describing several of these performance parameters can be highly misleading due to differences in definition and measurement techniques (1,8). For example, device sensitometry (linearity and sensitivity), resolution (modulation transfer function), Fourier plane signal-to-noise ration, uniformity (global vs. local), and exposure-dependent phase present particular complications in specification and interpretation. Sensitometry, resolution, and exposure-dependent phase depend on a wide range of parameters (including, for example, wavelength input, exposure magnitude, exposure pulse duration, operational mode (biasing conditions), readout mode, magnitude of baseline subtraction) since SLM's are truly active devices. Resolution (MTF) specifications also depend strongly on the measurement technique and analytical method employed (8) due to physical differences in device response to, for example, holographic fringe exposure as opposed to incoherent grating exposure at a given wavelength. Measurement of Fourier plane signal-to-noise ratio and device uniformity are largely dependent on defect densities that vary greatly from device to device due to lack of appropriate manufacturing process control.

The problems noted above with regard to accurate parametric- and application-dependent characterization of SLM's are by no means insurmountable. Rather, they point out the importance of continued advances in the fundamental physical understanding of device design and operation. As will be illustrated in several cases below, such improved understanding can indicate optimum design criteria, new operational modes, and new applications for traditional SLM's, as well as provide direction for the development of novel SLM technologies.

In the following three sections, current progress in each of three categories of spatial light modulators is highlighted, with empnasis on the physical principles of device

operation, principal advantages and disadvantages for potential applications, and where appropriate on current technological vs. fundamental limitations of device performance. The scope of this paper is necessarily limited, and the goals are to describe recent results that demonstrate the importance of advances in device understanding on SLM operation, as well as to point out a number of promising novel SLM technologies. This focus has necessitated some selectivity in the types of devices treated. Spatial light modulators based on thermoplastic, deformable oil film, deformable membrane, electrooptic ceramics (PLZT), magnetooptic, photochromic and photodichroic effects have been reviewed previously (see Refs. 1-7). In Section III, progress in electrooptic spatial light modulators is described, including the Pockels Readout Optical Modulator (PROM), the Microchannel Spatial Light Modulator (MSLM), and the photo- and electron beam-DKDP devices. Section IV covers recent advances in liquid crystal spatial light modulators, including the hybrid field effect liquid crystal light valve (LCLV), the Si-addressed LCLV, and the newly developed CCD-addressed LCLV. In addition, two novel liquid crystal devices are described for nonlinear processing applications: the multiple period LCLV for parallel A/D conversion and the variable grating mode liquid crystal device for implementation of arbitrary point nonlinearities and optical logic and computing applications. Photorefractive effect spatial light modulators for applications in correlation/convolution, edge enhancement, double-exposure holographic testing, and phase conjugate image generation are described in Section V.

III. Electrooptic Spatial Light Modulators

Electrooptic spatial light modulators (ESLM's) have been investigated for numerous applications in coherent optical signal processing, including the Pockels Readout Optical Modulator (PROM) (9), the Microchannel Spatial Light Modulator (MSLM) (10,11), the electron beam-DKDP SLM (TITUS) (12,13), and the photo-DKDP SLM (PHOTOTITUS) (14,15). Such devices record two-dimensional image information in the form of a charge pattern which modulates the voltage across an active (electrooptic) single crystal layer. The charge pattern is typically induced either by intensity modulation of light incident on a photoconductive layer, or by direct electron beam charge deposition. More recently, image storage in ESLM's has been accomplished by exposure to x-ray sources (13,16) and by high energy beam charge implantation (13,17). In each case, the image-wise modulated voltage is sensed using polarized light by means of the linear electrooptic effect.

Since all four of these electrooptic spatial light modulators have a number of similar operational features, the description of their essential characteristics will be simplified by examining the physical operation of the PROM first as an example, followed by discussions of each of the other devices.

III.1 Pockels Readout Optical Modulator (PROM)

A schematic diagram of a typical PROM structure is shown in Figure 1. The device is comprised of a photoconductive, electrooptic crystal wafer ($Bi_{12}SiO_{20}$) sandwiched between two thin dielectric blocking layers. The blocking layers are coated with transparent electrodes. The operation of the PROM is shown schematically in Figure 2. A voltage (typically 2 kV) is applied to the device, dividing between the dielectric and electrooptic layers in inverse proportion to the capacitance of each layer. A pulse of uniform uv illumination generates free electron-hole pairs which are subsequently separated by the applied field. Since the mobility-lifetime product of electrons is much larger than for holes in $Bi_{12}SiO_{20}$, illumination of the negative electrode provides the most efficient field reduction per absorbed photon. In this "erase/prime" step (Figure 2), sufficient uv illumination is provided to effectively cancel the internal field in the BSO layer. Note that the dielectric blocking layer thicknesses must be sufficient to witnstand 1/2 V_{app}. When the applied bias V_{app} is now reversed, the applied field and the field due to the stored (displaced) charges add rather than cancel, resulting in approximately $2V_{app}$ dropped across the BSO layer. Illumination from the negative electrode side with image-wise modulated blue light causes electron-hole pair generation and subsequent charge separation at a rate proportional to the incident intensity at each location, giving rise to a reduction in voltage across the electrooptic crystal in illuminated regions. Since in the absence of illumination (subsequent to the writing process) the dielectric relaxation time ($\varepsilon\varepsilon_0\rho$) of bismuth silicon oxide is unusually large (of order several thousand seconds), the electron distribution is trapped in the bulk; hence, the written charge distribution is stored. The resultant two-dimensional voltage distribution induces a birefringence (through the linear longitudinal electrooptic effect) which alters the polarization of linearly polarized readout light oriented to bisect the principal birefringent axes, producing an image-wise modulated amplitude when viewed through a crossed analyzer (see Figure 3). The readout amplitude transmitted through a PROM between ideal crossed polarizers may be expressed as

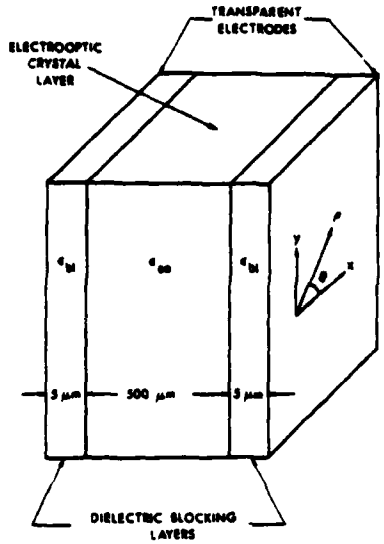


Figure 1. Structure of a typical PROM. In current PROM's, the electrooptic crystal layer is bismuth silicon oxide ($Bi_{12}SiO_{20}$, $\varepsilon_{eo}=56\varepsilon_0$) and the dielectric blocking layers are parylene C ($\varepsilon_{bl}=3\varepsilon_0$).

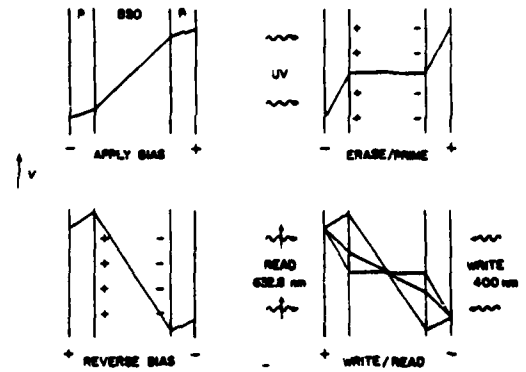
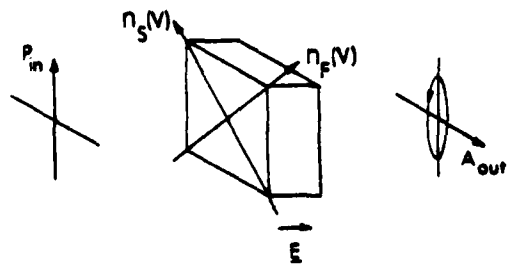


Figure 2. Diagram of basic PROM operational sequence, showing the potential distribution within the parylene and $Bi_{12}SiO_{20}$ layers following each step in the erase/prime and write/read sequence.



$$A_{out} = A_0 \sin\left(\frac{\pi V}{2 V_{\pi}}\right) \sim A_1 V \qquad \{V \ll V_{\pi}\}$$

Figure 3. Electrooptic spatial light modulator readout configuration through crossed polarizers.

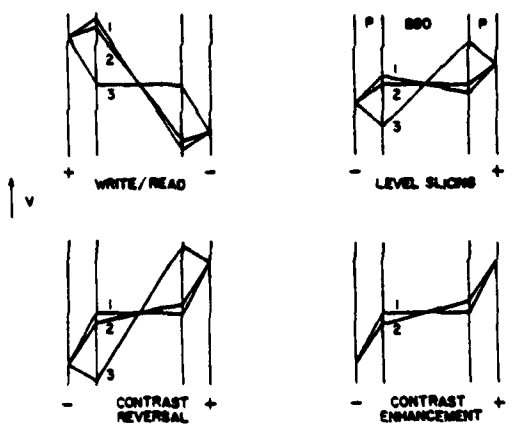


Figure 4. Diagram of several PROM optical processing operations. Shown are the potential distribution within the parylene and $Bi_{12}SiO_{20}$ layers for the negative write/read, level slicing, contrast reversal, and contrast enhancement functions.

$$A(x,y) = \frac{A_0}{\sqrt{2}} \sin\left(\frac{\pi V(x,y)}{2V_{\lambda/2}}\right) \qquad (1)$$

where A_0 is the incident readout light amplitude, $V_{\lambda/2}$ is the half-wave voltage of the electrooptic crystal, and $V(x,y)$ is the voltage across the electrooptic crystal at image coordinate (x,y). It should be noted that in the above expression it is assumed that the readout wavelength and intensity are chosen such that insignificant photoconductive charge redistribution occurs, and the effects of natural optical activity in the bismuth silicon oxide crystal may be neglected (18,19). Thus the output amplitude is a monotonic function of the input intensity (for applied voltages less than the electrooptic half-wave voltage); such a transfer relationship is desirable for incoherent-to-coherent conversion and subsequent coherent signal processing operations.

The write/read mode described above is presented again in Figure 4, where three different levels of exposure are depicted. Note that in this mode of operation a negative of the input image is produced on readout. An image positive may be created easily by reversing the sense of the applied voltage, creating a contrast reversal, as shown. In addition, contour generation (level slicing) is achievable by varying the external bias, causing various internal field regions to be cancelled by the external applied field. In the illustration chosen in Figure 4, the region labeled "2" has been brought to the null condition by suitable reduction of the applied voltage. Contrast enhancement of highly underexposed images results from external bias adjustment to disperse the exposed regions about the zero field condition, as shown.

Recent progress has been achieved in the analysis of the fundamental resolution limitations of the PROM as well as of the other electrooptic spatial light modulators (20-22). From the nature of Eq. (1), it can be seen that the resolution of an electrooptic spatial light modulator depends directly on the relationship between a periodic (spatial frequency ω) variation in the writing intensity, and the resultant spatial modulation of the voltage across the electrooptic crystal. The cubic symmetry (I23) of bismuth silicon oxide in conjunction with the orientation of the electrooptic crystal slice (<001>) assures that only longitudinal components of the electric field contribute to the induced birefringence through the electrooptic effect. Therefore, the modulation in output amplitude [see Eq. (1)] depends only on the voltage difference $V(x,y)$ between opposite sides of the electrooptic crystal at each image point (x,y). In the absence of significant two-dimensional diffusion effects in comparison with the drift-aided charge separation of photogenerated electron-hole pairs, $V(x,y)$ will be a function of the dielectric constants and thicknesses of the electrooptic and blocking layers, the locations (in the z-direction) of the trapped hole and electron distributions resulting from the writing (image recording) process, and the spatial frequency of the charge (writing intensity) modulation in the (x,y) plane.

From an exact solution of the device multi-layer structure containing a single point charge, the Fourier transform of the voltage distribution (which can be directly related to the modulation transfer function) was derived as a function of the charge location within the electrooptic crystal and the device layer parameters (20). Since the modulation transfer function (MTF) for a given device will be both exposure and modulation dependent, a more fundamental indication of expected device performance is obtained by presentation and discussion of the un-normalized $[V(\omega)]$ and normalized $[V(\omega)/V(0)]$ potential difference functions.

At this point, it should be mentioned that the un-normalized potential differences $V(\omega)$ and the normalized function $V(\omega)/V(0)$ have a direct interpretation useful for both comparison of distinct devices of different constitutive characteristics, and for comparison of the implications of distinct charge distributions (resulting from different exposure parameters) within a given device. In particular, graphs of $V(\omega)$ as a function of ω parameterized by different device constitutive properties assume equal exposure conditions (identical charge distributions) while graphs of $V(\omega)/V(0)$ assume optimum exposure conditions for each compared device (i.e., sufficient exposure for each device such that $V(\omega)$ is optimized in the limit of low spatial frequencies. These types of comparisons are familiar from the case of photographic film (23) where the typical resolution/sensitivity trade-off forces a similar comparison of film properties on the basis of either response to equal exposure, or response to optimum exposure (see Figure 5).

This situation is graphically illustrated by reference to Figures 6 and 7, which depict the spatial frequency dependence of $V(\omega)$ and $V(\omega)/V(0)$ for PROM device parameters as shown in Figure 7, with the dielectric constant of the dielectric blocking layer as a parameter. The effect of increasing the blocking layer dielectric constant is seen to reduce the device sensitivity (see Figure 6), while increasing the high spatial frequency response for optimum exposure in both cases (see Figure 7). In Figures 6 and 7, a single point charge is located

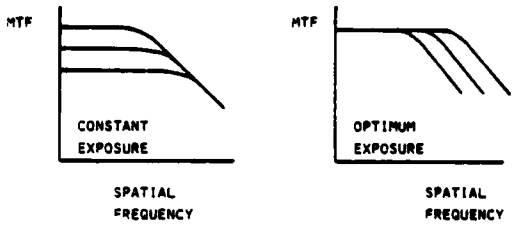


Figure 5. Resolution/sensitivity tradoff for photographic film. The unnormalized curve shows the reduction in sensitivity that accompanies increased resolution capability.

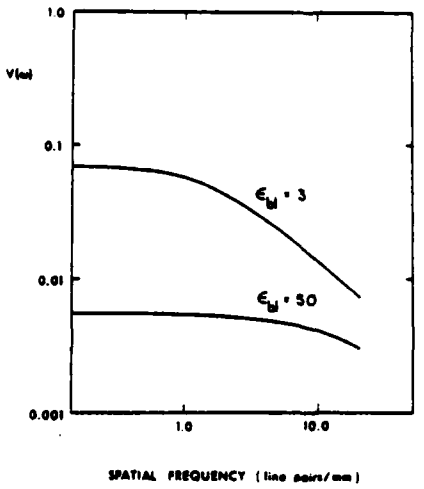


Figure 6. Effect of dielectric blocking layer dielectric constant on the spatial frequency response of V(ω). A symmetric PROM device is assumed, with geometric and constitutive parameters as shown in Figure 1.

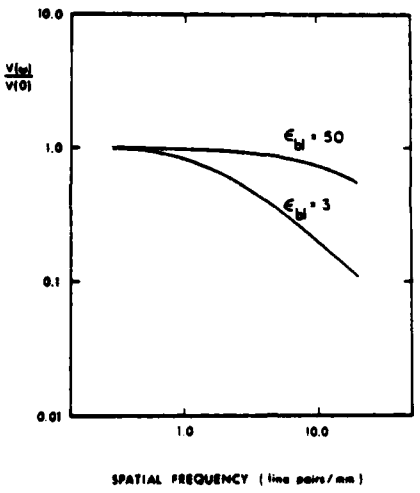


Figure 7. Effect of dielectric blocking layer dielectric constant on the spatial frequency response of V(ω)/V(0).

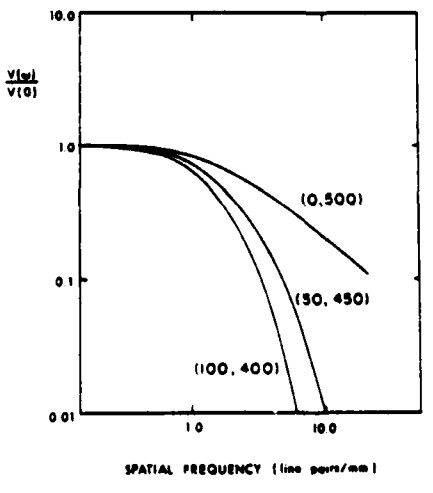


Figure 8. Effect of longitudinal charge position witnin the electrooptic crystal layer on the spatial frequency response of V(ω)/V(0). Note the marked attenuation of the high spatial frequency response as the charges are increasingly displaced from their respective interfaces.

at one of the dielectric blocking layer/electrooptic crystal interfaces.

The dependence of the spatial frequency response on the dielectric constants and thicknesses of the dielectric blocking layers (as shown in Figures 6 and 7, and in Figures 5 and 6 of Ref. (20)) indicates that new PROM devices can be envisioned with modulation transfer functions constant to much higher spatial frequencies than are characteristic of presently available devices. Gains in MTF behavior due to choice of dielectric blocking layer properties will be accompanied by an overall reduction in device sensitivity. Such new devices, however, would provide significantly improved image fidelity and resolution in applications where requirements on device sensitivity can be relaxed (as is the case with photographic film). In order to improve PROM resolution, high dielectric constant, high dielectric breakdown strength, high resistivity dielectric blocking layers are required.

In an effort to understand the effects of exposure-induced charge distributions throughout the bulk of the electrooptic crystal layer on PROM resolution, solutions for cases involving multiple point charges were obtained by a linear superposition of the solutions for each separate point charge (21). In particular, the effect of an electron-hole pair can be modeled if the point charges are assigned opposite signs. The results of such a calculation as a function of the charge separation are shown in Figure 8. In this calculation, the hole and electron were assumed initially constrained to opposite interfaces, and were subsequently displaced symmetrically into the bulk of the electrooptic crystal layer. Three cases are depicted in Figure 8, in which the hole and electron were first constrained to opposite dielectric blocking layer/electrooptic crystal interfaces (0,500), and were subsequently displaced symmetrically into the electrooptic crystal layer by 50 μm (50,450) and 100 μm (100,400), respectively. From Figure 8, it is observed that charges displaced from the dielectric blocking layer/electrooptic crystal interfaces strongly degrade the high spatial frequency response. In addition, such displacement reduces the sensitivity at low spatial frequencies (21).

An estimate of the device spatial frequency response resulting from optical exposure may be obtained from iterative solutions of the charge transport equations under simulated exposure conditions (24,21). Sample charge distribution results corresponding to exposure wavelengths of 450 nm and 375 nm for device constitutive properties as described in Figure 1 are shown in Figure 9. The absorption coefficients (α= 30 cm^{-1} @ λ = 450 nm; α = 830 cm^{-1} @

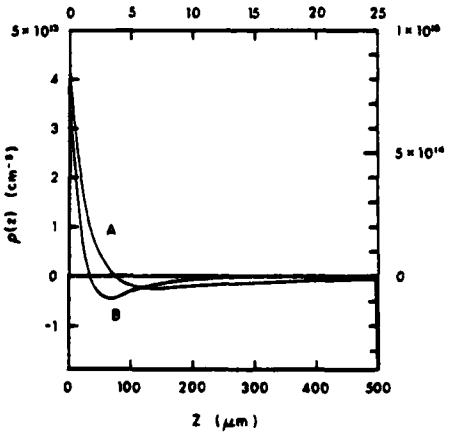


Figure 9. Longitudinal charge distribution $\rho(z)$ calculated numerically using an iterated electron drift model for two exposure wavelengths (and corresponding absorption coefficients). Curves A and B result from exposure wavelengths of 450 nm and 375 nm, respectively. The left and lower scales correspond to curve A, while the right and upper scales correspond to curve B.

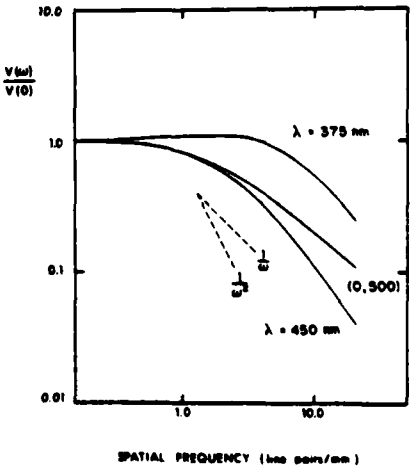


Figure 10. Normalized frequency responses $V(\omega)/V(0)$ for the calculated longitudinal charge distributions of Figure 9. The normalized response of point charges at the two interfaces is included for comparison.

λ = 375 nm) are such that the 375 nm case is characterized by close confinement of the hole distribution near the input interface with shallow electron penetration into the bulk. In contrast, the 450 nm case is characterized by significant electron-hole pair generation throughout the electrooptic crystal layer, which results in a sizeable accumulation of electrons at the far interface (as shown in Figure 9).

The spatial frequency responses calculated from these widely disparate charge distributions are shown in Figure 10. The closely confined charge in the 375 nm case produces a nonmonotonic spatial frequency response (20-22), whereas the volume excitation present in the 450 nm case generates a monotonic frequency response. In both cases, the high spatial frequency response asymptotically approaches ω^{-2}, in contrast to the ω^{-1} asymptote applicable to cases where all charges are confined to the interface (21).

The effect of exposure-induced charge distribution throughout the bulk of the electrooptic crystal layer is seen to have a marked effect on the high spatial frequency response characteristics of the image storage process. The optimum resolution and sensitivity within a given PROM device structure are obtained when the hole distribution is constrained to the interface nearest the negative electrode, and the entire electron distribution is swept to the interface nearest the positive electrode during the writing cycle. The actual resolution and sensitivity obtained for a given device are thus strongly dependent on the absorption coefficient of the electrooptic layer at the writing illumination wavelength, on the external applied voltage and voltage division between the multiple layers during the writing cycle, and on the mobility-lifetime product of photogenerated electrons in the electrooptic crystal. Such bulk charge distribution effects will also strongly affect the resolution and sensitivity for cases where the image-wise modulated charge pattern is induced by high energy electron beam (13,17) and x-ray (16 sources. Specification of the exposure wavelength (in addition to exposure level, bias exposure, operational mode, and MTF test method (8)) is thus seen to be critical for proper device evaluation and comparison. Research on the physics of the device operational modes (prime, erase, exposure, and "superprime" (9)) utilizing the charge transport and resolution models described above is currently in progress.

The PROM will continue to be utilized in applications requiring either temporal storage or time integration. Current PROM's exhibit excellent contrast and Fourier plane signal-to-noise ratio, moderate sensitivity and resolution, and provide a number of active image preprocessing functions. The principal shortcomings of the PROM relative to certain applications are nonlinear sensitometry effects at high exposure levels, the necessity of blue write wavelengths for optimum sensitivity and resolution, and the lack of a completely nondestructive readout mode (which limits the available readout gain).

III.2 Microchannel Spatial Light Modulator (MSLM)

The Microchannel Spatial Light Modulator (MSLM) (10,11) is under development for coherent optical processing applications requiring high sensitivity, such as stellar speckle interferometry or low-visibility optical communication. The MSLM is shown schematically in Figure 11, consisting primarily of an evacuated cell (which may be either sealed or demountable), a photocathode, a microchannel array plate, a dielectric mirror, and a thin electrooptic crystal layer. The microchannel array plate consists of an array of semiconducting glass-lined pores (~10 μm diameter), and is bounded by two semi-transparent electrodes. The pores are oriented at an angle to the plate normal, so that incident electrons impact the pore walls, giving rise to electron multiplication by successive secondary electron emission. The microchannel plate is separated from the electrooptic crystal by a gap (~500-1000 μm). In combination with the dielectric multilayer mirror, this gap serves a voltage division function similar to the dielectric blocking layer in the PROM.

In operation, an initial electron distribution is emitted from the photocathode in response to an incoherent or coherent control image (within the wavelength region of photosensitivity of the photocathode). After electron multiplication in the appropriately biased microchannel array plate, the amplified electron image is proximity focused onto the dielectric mirror. Either positive or negative charge distributions can be written on the mirror surface, depending on whether the ratio of secondary (emitted) electrodes to primary (incident) electrons is greater or less than unity (see Figure 12). The charge distribution induces a spatially varying longitudinal electric field field in the electrooptic crystal layer, which modulates the local refractive indices through the Pockels effect. As in the case of the PROM, either phase or amplitude modulation can be achieved on reflective readout by appropriate readout polarization and/or orientation of the electrooptic crystal. In current devices (11), both $LiNbO_3$ and $LiTaO_3$ z-cut plates have been utilized to produce pure phase modulation.

The dynamic operation of the MSLM can be understood with reference to Figure 13, which depicts lines of stable equilibria (full bold lines) and a line of unstable equilibria (broken bold line) in the state-space of the device (gap energy eV_g as a function of

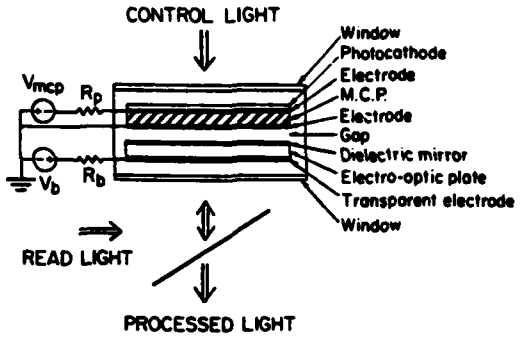


Figure 11. The microchannel spatial light modulator (MSLM): proximity-focussed configuration. (After Warde, Ref. 10).

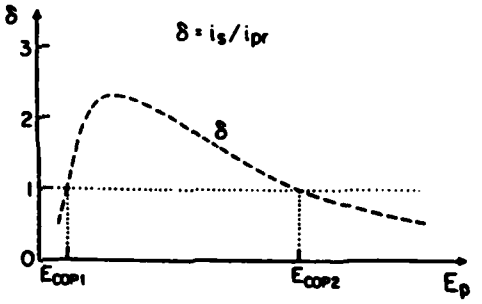


Figure 12. Secondary electron emission characteristics of a typical insulator. (After Warde, Ref. 11).

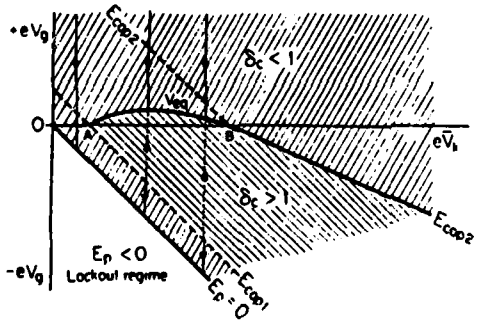


Figure 13. State-space plot illustrating the dynamic operation of the MSLM. (After Warde, Ref. 11).

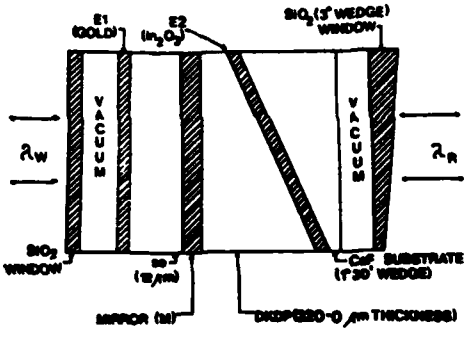


Figure 14. Schematic diagram of the photo-DKDP spatial light modulator. (After Casasent, Ref. 14).

effective cathode potential $e\bar{V}_k$). The effective cathode potential is a function of the secondary emission characteristics of the dielectric mirror/crystal surface, the gap voltage V_g, the microchannel plate voltage V_{mcp}, and the shape of the electron energy distribution emitted from the microchannel plate. The parameter δ_c is the ratio of secondary electrons collected by other parts of the system to primary electrons. With no illumination incident on the photocathode, the gap voltage may be altered by raising or lowering V_b (see Figure 11). If the gap voltage is initially biased into the region where $\delta_c<1$, illumination of a particular region of the photocathode will induce electron deposition on the crystal, causing the local gap voltage to drop along one of the vertical lines shown in Figure 13. Saturation in this region will occur when the total local exposure exceeds the level required to drive the gap voltage onto the stable equilibrium curve. The image can be erased by subsequently lowering the bias into the region $\delta_c>1$ and uniformly or selectively illuminating the photocathode, such that net electrons are emitted from the surface, raising the local gap voltage along one of the dashed lines shown. The lower dashed bold line labeled E_{cop1} represents unstable equilibria, since lowering of the gap voltage below this line causes entry to a second $\delta_c<1$ region, such that illumination drives the gap voltage toward the line $E_p = 0$ along which the incident primary electron energy at the crystal surface is zero. Hence once a local region of the crystal surface resides on this line, the local potential cannot be further altered without changing the bias level V_b.

Charge distributions have been successfully stored directly on a 500 μm thick $LiTaO_3$ crystal (without a dielectric mirror) for periods as long as two weeks (11). Although the surface and bulk resistivities of $LiTaO_3$ are such as to result in large dielectric relaxation time constants, it is likely that deep surface trap states with extraordinarily long relaxation times must be involved in the electron-induced charge storage mechanism. Research on the nature of such trap levels by spectroscopically selective photoemission is currently in progress (25). The existence of image storage capability can be used in concert with programmable variation of the bias voltage V_b to allow implementation of several useful image preprocessing functions. Image addition and subtraction can be performed by temporal integration and sequential selective write/erasure, respectively. Contrast enhancement and reversal can be implemented by bias adjustment following exposure, as described earlier in the case of the PROM. Utilization of the erasure saturation characteristic described above provides a form of variable level thresholding. In addition, edge enhancement is possible due to variation of the secondary electron emission characteristics in the presence of fringing fields in a region of exposure discontinuity.

One of the most promising features of the MSLM configuration is its inherent flexibility in both choice of electrooptic crystal and choice of photocathode material (which together determine the readout and writing wavelength response regions for the device, respectively). Since the microchannel array plate is inherently responsive from x-ray to near-uv wavelengths, no photocathode is required for operation in this regime. The device exhibits extremely high sensitivity, and the physical separation of the write and read functions allows further inherent throughput gain in image amplifier applications. The combination of long term storage with (at present) near-TV frame rates allows great flexibility in system design. Current MSLM's exhibit resolutions (~3 lp/mm) limited by the thickness of the electrooptic crystal, the relatively thick dielectric blocking layer (gap), and the low dielectric constant of the gap. Although the implementation of the active image preprocessing functions described above involves rapid programmability of relatively high voltages, the almost purely capacitive loading presented by the MSLM should minimize such difficulties.

III.3 Photo-DKDP Spatial Light Modulator

The photo-DKDP Spatial Light Modulator (also called "PHOTOTITUS" (15)) has been developed and is currently manufactured by Laboratoires d'Electronique et de Physique Appliquee in France, and has only recently become available for experimentation in the United States (26). A schematic diagram of the photo-DKDP device is shown in Figure 14. A wedged electrooptic crystal layer of deuterated potassium dihydrogen phosphate (DKDP) is sandwiched between a dielectric mirror and a transparent electrode on a CaF substrate. A photoconductive layer of amorphous selenium is deposited on the dielectric mirror, and is overcoated with a second transparent electrode. This assembly is incorporated in an evacuated two-stage Peltier cooler to allow operation near the ferroelectric Curie point of DKDP (~-50°C), which enhances the resolution and storage time of the device while significantly reducing the half-wave voltage of DKDP to approximately 300V. This reduction in operating voltage is critical in order to allow the incorporation of the Se photoconductor, which effectively separates the read and write functions.

Operation of the photo-DKDP SLM is similar in most respects to the PROM operational modes described above, as is depicted schematically in Figure 15. In Figure 15(a), initial application of a positive voltage V_0 to the Se electrode produces a voltage division among the various layers. Exposure to appropriate wavelength write illumination (typically 440-520 nm) induces hole transport across the Se layer, reducing the voltage across the

photoconductor and increasing the voltage across the electrooptic DKDP crystal (Figure 15(b)). In Figure 15(c), the electrodes have been short-circuited to produce a positive readout image. The small residual voltage V_2 in the dark (unilluminated) region arises from the finite dark conductivity of the photoconductive layer during the writing cycle. Erasure is accomplished by subsequent uniform illumination of the photoconductor to induce electron transport to the mirror interface, thereby reducing the stored electrostatic field to zero. It should be noted that exposure wavelengths on the long wavelength side of the selenium photoconductivity response will initiate bulk electron-hole pair generation with resultant ambipolar diffusion. Since the mobility-lifetime product for electrons is significantly smaller than that for holes, sensitometry curves resulting from primarily electron transport differ from those characterized by primarily hole transport. These differences are especially important for image subtraction applications (14).

The photo-DKDP spatial light modulator has been utilized in a wide range of coherent optical data processing applications, and has demonstrated processing accuracies comparable to those achievable with photographic film inputs (26). The device as presently configured has moderate sensitivity, storage capability, and the availability of image subtraction in addition to other image preprocessing functions. The resolution is enhanced relative to that exhibited by current PROM's due to the increase in dielectric anisotropy resulting from Curie point operation. On the other hand, this very feature limits the contrast available in non-collimated readout configurations due to large natural birefringence. The necessity of cooling to achieve proper device performance can led to nonuniformities in device characteristics, since small temperature differences yield large changes in the magnitude and ratio of the dielectric constants. The hygroscopic nature of DKDP necessitates the use of an evacuated cell; however, even with this constraint the reported optical quality of current devices is excellent (14).

III.4 Electron-Beam DKDP Spatial Light Modulator

The electron-beam DKDP SLM has been under development for many years for coherent optical processing applications, and has been implemented during the past ten years in two similar versions (12,13). The primary difference between this electrooptic spatial light modulator and the three previously described is the mode of address. As shown in Figure 16,

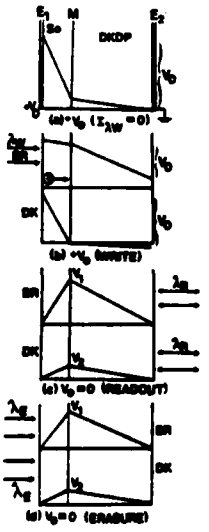


Figure 15. Distribution of voltage as a function of distance in the photo-DKDP SLM at different operational steps for illuminated (BR) and unilluminated (DK) regions. (After Casasent, Ref. 14).

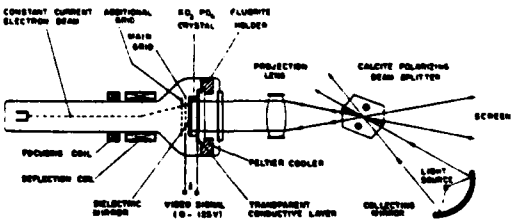


Figure 16. Electron-beam-addressed Pockels effect imaging device using a DKDP (KD_2PO_4) crystal plate operated in the reflection mode. A constant current electron beam is used and the video signal is applied between the transparent conductive layer and the main grid placed at about 40 μm from the target. (After Marie, Ref. 13).

charge is deposited on a dielectric mirror/electrooptic crystal layer by an axial (13) or off-axis(12) electron gun driven in a raster pattern and grid-modulated by a serial (usually video) input. The physics of operation of this device thus combines several features of the MSLM and the photo-DKDP SLM. The design, operation, and applications of this SLM have been extensively documented (12,13) and are not treated herein. In addition to a wide range of applications in optical data processing, the electron-beam DKDP spatial light modulator has been successfully utilized for large-screen multi-color television projection display (13).

IV. Liquid Crystal Spatial Light Modulators

A number of spatial light modulator technologies are based on the utilization of liquid crystals. The attractiveness of liquid crystal layers for this application arises primarily from several unusual characteristics of nematic liquid crystals, including large dielectric anisotropies (resulting in low operating voltages), large birefringences (resulting in sizeable polarization effects even in the thin cells necessary for rapid response times), and a wide variety of possible alignment configurations. In this section, several liquid crystal spatial light modulators are discussed, including some potentially exciting modifications of the Hughes liquid crystal light valve, as well as two novel liquid crystal devices for nonlinear coherent optical processing applications.

IV.1 Hybrid Field Effect Liquid Crystal Light Valve

By far the most advanced spatial light modulator employing a liquid crystal layer as the active electrooptic element is the hybrid field effect liquid crystal light valve (LCLV) (27,28). The typical device configuration of the LCLV is presented in Figure 17. A transparent electrode and a chemically inert insulating layer (SiO_2) are deposited on an optically flat glass substrate. The insulating SiO_2 layer prevents dc current flow through the device while simultaneously functioning as a preferred direction alignment layer for the liquid crystal and as an ionic blocking layer to prevent cell poisoning effects. A biphenyl nematic liquid crystal layer is confined laterally by a deposited thin film spacer, and longitudinally by a second SiO_2 alignment layer. The liquid crystal layer is photoconductively-addressed by a thin film of cadmium sulfide deposited on a transparent electrode-coated optically flat glass substrate. Separation of the read and write functions is accomplished by incorporation of a dielectric mirror and cadmium telluride light blocking layer between the liquid crystal layer and the CdS photoconductor, as shown.

In order to optimize the electrooptic properties of the liquid crystal layer for coherent optical processing applications, a hybrid field effect operational configuration is employed (27). In this configuration, the uniaxial liquid crystal molecules are preferentially aligned (by means of appropriate surface treatment of the SiO_2 alignment layers) with their long axes parallel to the electrode surfaces (homogeneous alignment). In addition, the alignment directions on the two opposed surfaces are oriented to form an included angle of 45°. This 45° twist imparts a continuous rotation of the liquid crystal molecules from one surface to the other, as shown schematically in Figure 18. Since the nematic liquid crystal employed in this device exhibits large birefringence (difference in refractive indices for light polarized parallel and perpendicular to the molecular long axis), the resultant twisted layer can be modeled as a succession of thin birefringent layers, each oriented at a slight angle with respect to the immediately preceding and following layers. Such an optical configuration has been shown to be equivalent under certain conditions to a purely optically active layer (29), so that the polarization of light traversing the layer undergoes a pure rotation of 45°.

The operation of the hybrid field effect liquid crystal layer can be explained with reference to Figure 19. In the "off" state with no voltage applied across the liquid crystal layer, input light polarized parallel to the preferential alignment direction of the entrance electrode experiences a positive 45° polarization rotation on traversing the layer, and a negative 45° rotation following reflection from the dielectric mirror and a second pass through the cell. Hence the emergent polarization is parallel to the incident polarization, and can be extinguished with a crossed analyzer to provide a dark off state. When voltage is applied across the liquid crystal layer, the resulting longitudinal electric field tends to align the liquid crystal molecules with the field direction (due to the positive dielectric anisotropy of the molecules). This "tilt" toward perpendicular alignment from parallel alignment varies continuously between the two surfaces, and is largest in the center of the layer where the surface alignment forces are weakest. Due to the physical nature of the intermolecular forces, as the tilt angle increases in a given layer, the transmittance of the twist angle is weakened. Hence as the voltage is increased, molecules in the layer center approach a perpendicular orientation, allowing molecules near each surface to relax toward an untwisted state with orientation parallel to the induced surface orientation. The polarization of incident light will thus be relatively unaltered by the front half of the cell on the first pass, but will encounter a birefringent layer oriented at 45° in the rear half of the cell, causing the polarization of light striking the dielectric mirror to be elliptical. Since birefringence adds on reflection (rather than cancels as in the case of optical activity), the emergent polarization from the cell will in

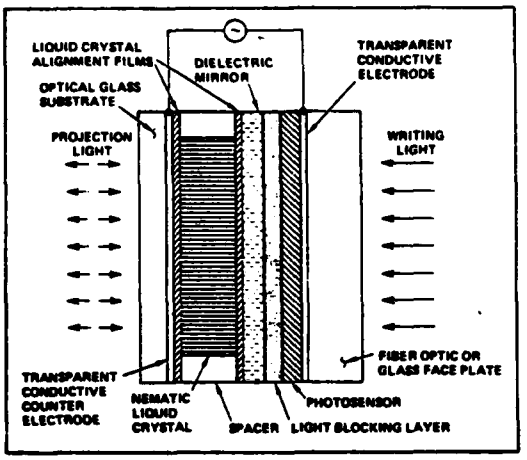


Figure 17. Cross-sectional schematic of the hybrid field effect liquid crystal light valve. (After Bleha, Ref. 27).

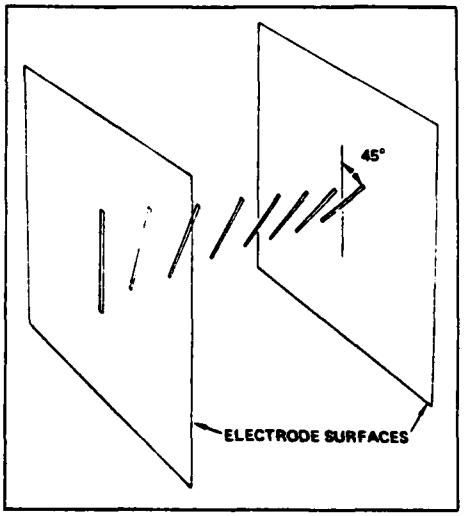


Figure 18. Schematic of 45° twisted nematic liquid crystal alignment (hybrid field effect model). (After Bleha, Ref. 27).

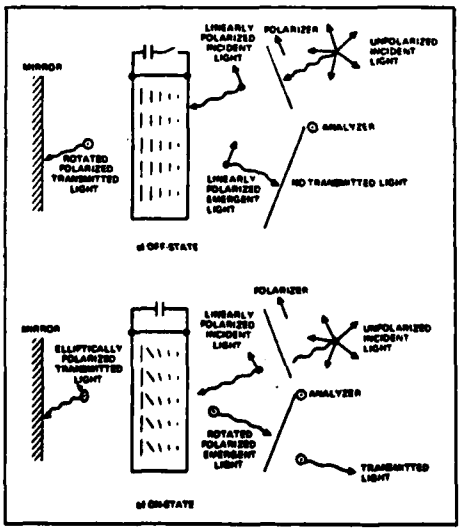


Figure 19. Operation of the hybrid field effect liquid crystal light valve: (a) the off-state; (b) the on-state. (After Bleha, Ref. 27).

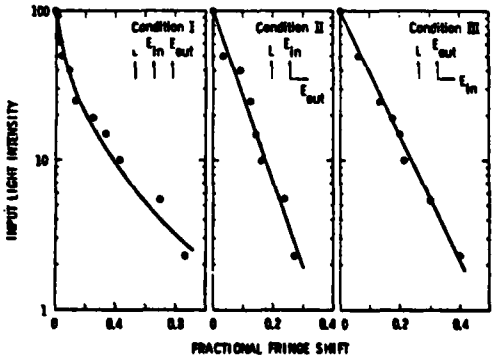


Figure 20. The phase response (in multiples of 2π rad) of the hybrid field effect LCLV as a function of input irradiance. The arrows indicate the relative configuration of the polarization of the coherent beam entering the device (E_{in}), the molecular orientation at this point (L), and the orientation of the analyzer (E_{out}). (After Gara, Ref. 37).

general be elliptical, such that the appropriate component will be passed by the crossed analyzer. A detailed analysis of the cell transmission as a function of voltage has recently been performed (30-32).

In order to function as an optical-to-optical converter, the LCLV requires a mechanism for spatially-dependent variation of the liquid crystal layer voltage in response to an input image intensity distribution. This function is performed by the CdS/CdTe heterojunction in series with the capacitance of the multilayer dielectric mirror. Analysis of the operation of this structure (33,34) shows that photocapacitive as well as photoconductive processes must be considered to fully explain the observed LCLV sensitometry behavior. The LCLV requires low operating voltages (5-10 V rms) at intermediate frequencies (1-10 kHz).

The principal advantages of the hybrid field effect liquid crystal light valve for coherent optical processing applications are good sensitivity and resolution characteristics, full separation of the write/read functions (which allows for the possibility of single wavelength optical feedback, for example), simplicity of operation, and low power requirements. The sensitometry characteristics of the LCLV depend rather strongly on the magnitude and frequency of the applied voltage (35,36); this feature can be used to advantage in some applications, but also contributes to device nonuniformity effects (36). Device uniformity in both the off and on states has been shown to be a quite sensitive function of liquid crystal layer thickness (30-32), which is particularly difficult to control in such a multilayer structure. Device reproducibility (both within a given device and from device to device) is dependent primarily on the deposition characteristics of the cadmium sulfide and cadmium telluride layers, since the optical and electronic properties of II-VI compounds are strong functions of deposition conditions. The device response time varies as a function of input intensity (primarily due to the characteristics of the CdS photoconductor), which when coupled with the lack of storage capability leads to temporal variations of the diffracted orders in the Fourier plane during operation, and in some cases to image ghosting with slow decay time constants in brightly exposed regions of the images. A number of these difficulties are due primarily to the utilization of CdS as the photoconductor, and should be eliminated with the advent of the Si-LCLV described below. Finally, it is important to note that the hybrid field effect structure gives rise to an input-intensity-dependent phase shift of the output wavefront that is a further function of the relative orientation of the LCLV, polarizer, and analyzer (37).

IV.2 Silicon Liquid Crystal Light Valve

Recently, a major potential technological advance in liquid crystal light valve technology has been reported (38), resulting from incorporation of a silicon photoconductor in a hybrid field effect device. The structure of this novel light valve is shown in Figure 21. The active liquid crystal layer is the same as that described previously for the CdS-addressed LCLV. The CdTe layer has been replaced with a cermet light blocking layer, composed of a multilayer structure of metallic (tin) islands dispersed in insulating (SiO_x) layers. A silicon dioxide gate insulator is coupled to a very high resistivity π-silicon wafer to form an MOS structure. A thin degenerately doped p^{++} layer forms the rear device electrode, and is overcoated with a thermally grown protective oxide coating. A square p-doped grid just beneath the SiO_2 gate insulator serves to enhance the device resolution (as described below), and p^+ isolation channel stops surround the periphery of the wafer to minimize minority carrier injection into the active region.

When the Si/SiO_2 interface is biased into accumulation, recombination at the interface erases any residual image and readies the device for the active mode. The Si/SiO_2 interface is then biased into depletion (depleting not only the p-grid but also the entire π-silicon layer), and the π-silicon side is illuminated with the input image. Illumination produces local electron-hole pair generation near the back contact, with electrons subsequently swept to the Si/SiO_2 interface by the field across the π-silicon layer. Calculations show there to be only negligible spreading of the swept-electron distribution during traversal across the wafer (39). In addition, the depleted p-grid is more negative than the surrounding π-regions, which acts to focus the incoming electrons into the π-buckets near the Si/SiO_2 interface. The p-grid also prevents significant charge spreading during the readout cycle. Due to the sequential accumulation/depletion cycle, the mode of operation is fully ac, which significantly extends the lifetime of the liquid crystal layer. Once the image-induced charge pattern has been established at the Si/SiO_2 interface, the electric field across the liquid crystal layer will be modulated such that the device may be read out in reflection just as described above for the CdS/CdTe LCLV.

Current performance parameters of the Si/LCLV are quite encouraging, with limiting resolution of 25 lp/mm at present limited by a 20 μm period grid, 20 $\mu W/cm^2$ sensitivity, and a contrast ratio of 30:1 (39). The device has been successfully operated at TV frame rates. With the incorporation of a grid with 6 μm period, the limiting resolution should be

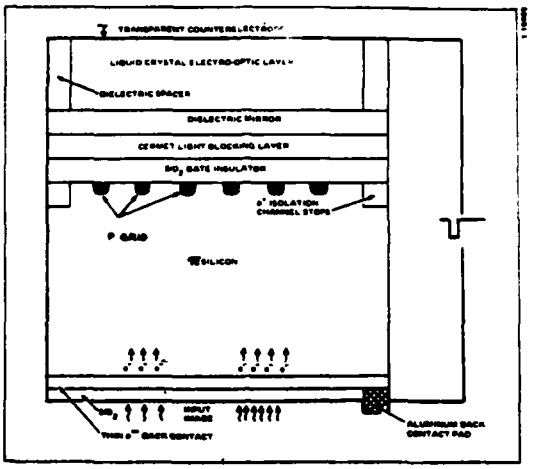


Figure 21. Schematic diagram of the photo-activated silicon liquid crystal light valve. (After Braatz, Ref. 38).

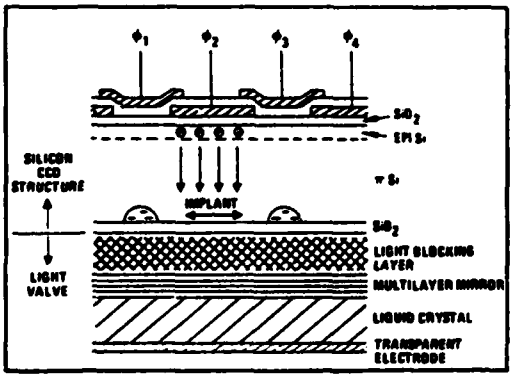


Figure 22. Cross-sectional schematic of the CCD-addressed liquid crystal light valve. (After Grinberg, Ref. 40).

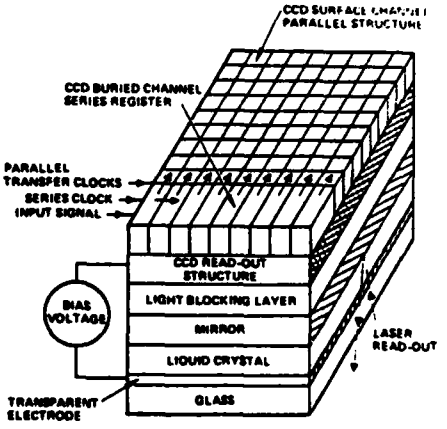


Figure 23. Functional schematic of the CCD-addressed liquid crystal light valve. (After Grinberg, Ref. 40).

extendable to 150 lp/mm. The liquid crystal layer is expected to accommodate such high spatial frequencies, since the presence of grid lines 3 μm wide has been observed in several test structures (39). The silicon photosensitivity spectrum extends from the near uv (400 nm) into the near ir (1.1 μm) with high quantum efficiency. Hence such Si-LCLV's should find numerous applications in optical data processing, including real-life scene input situations.

IV.3 CCD-Addressed Liquid Crystal Light Valve

Much of the new technology evident in the Si-LCLV is being utilized to mate a high bandwidth serial/parallel CCD register to a modified Si-LCLV for optical processing applications requiring a serial input format (40). As illustrated in Figure 22, the CCD-LCLV is similar in configuration to the Si-LCLV described above. The difference between the two consists primarily in the placement on the uppermost π-silicon surface of a surface channel CCD array. Proper sequencing of the clock lines ϕ_1-ϕ_4 propagates electron packets under each gate. Subsequent positive bias of the liquid crystal electrode drives the electron packets to the Si/SiO_2 interface, where they are focused and prevented from lateral diffusion by the implanted p-grid. A functional schematic diagram of the CCD array is shown in Figure 23. The input is serially transferred by the series clock into the CCD buried channel series register until an entire line of data is recorded. The line of data is then transferred in parallel to the adjacent CCD surface channel parallel structure as shown. When all CCD registers are full, representing a complete frame of input information, the bias voltage is changed to drive the entire stored electron packet array to the Si/SiO_2 interface. At this point, the two-dimensional image can be read out in reflection by polarized illumination from the liquid crystal side of the device, as shown. The buried channel serial array is necessitated by the 100 MHz design clock frequency, while the slower (100 kHZ) parallel array requires surface channel technology to allow for subsequent charge transfer to the Si/SiO_2 interface.

Current CCD-LCLV devices are 64x64 arrays of elements on 1.3 mil centers (total active area 83 mils square) (39). A 256x256 element array is presently under development, with an eventual goal of a 1000x1000 element array operating at 100 kHz per line. Such a device would operate at an overall frame rate approaching 100 HZ. Several difficult technical problems currently under investigation include the quality of image transfer from the CCD to the LC layer, development of the novel buried channel serial/surface channel parallel CCD approach, uniformity of processing of the very high resistivity π-silicon wafer, and the processing delicacy required for the 5 mil thinned wafers. Achievement of these design goals will represent a long-awaited major breakthrough in serial input/parallel output spatial light modulator technology.

IV.4 Multiple Period Liquid Crystal Light Valve

A technique for optically performing parallel analog-to-digital conversion on incoherent two-dimensional inputs at real-time rates utilizing a multiple period liquid crystal light valve has been recently described (41). This MP LCLV is similar in construction to the hybrid field effect LCLV with the notable exception that the liquid crystal layer was homeotropically aligned (long molecular axes perpendicular to the electrode in the "off" state). The liquid crystal chosen has negative dielectric anisotropy, so that application of a voltage across the layer tends to rotate the molecules parallel to the electrodes. The magnitude of the rotation is a function of the applied voltage, so that readout light polarized at 45° with respect to the projection of the long molecular axis on the electrode surface (in the partially rotated state) will experience pure birefringence on traversal of the cell. For liquid crystals with large optical anisotropies (and for thick enough cells), the total phase retardation can be many multiples of 2π. Since the local voltage across the liquid crystal layer can be photoconductively-addressed, the overall relationship between the intensity transmittance of the device and the incident intensity at any point is given ideally by the sinusoidal ($\sin^2$) curve shown by a dashed line in Figure 24. In Figure 24(a), optical or electronic thresholding of the device transmittance at one half produces the least significant bit of the reflected or Gray code. Rescaling the input intensity by a factor of one half produces the device transmittance curve shown in Figure 24(b). Thresholding of this curve at one half produces the next most significant bit, and so on. Sequential rescaling as shown in Figure 24(a-c) produces the 3 least significant Gray-code bit planes in sequence. Parallel rescaling is also possible utilizing 3 periodically repeated attenuating strips on the device surface of attenuation factors 1, 1/2, and 1/4, respectively (41).

Due to the nonlinear sensitometry effects associated with the cadmium sulfide photoconductor described in Section IV.1, the actual device response curve departs significantly from the ideal behavior of Figure 24. The response curve of the actual device utilized for the A/D conversion experiments is shown in Figure 25. The aperiodic nature of the device response curve necessitated nonuniform quantization level assignments, as shown in the figure. Nevertheless, real time parallel three-bit A/D conversion was performed successfully with this device (41), with an estimated potential A/D conversion rate of

1.2×10^8 points per second for currently available device parameters. Development of a LCLV with a more nearly periodic response function would significantly advance the potential of this technique, but requires linearization of both the photoconductor response and the relationship between applied voltage and effective birefringence.

IV.5 Variable Grating Mode Liquid Crystal Device

Variable grating mode (VGM) liquid crystal devices offer a new approach to the problem of optical transducers (42-44) for nonlinear optical processing and optical logic and computing applications. The basic function of the VGM device is to perform an intensity-to-spatial frequency conversion over a two-dimensional image field. In this process, the intensity variations of an input image are converted to local spatial frequency variations in a phase grating structure within the liquid crystal layer. Due to this intensity-to-spatial frequency conversion, a standard spatial filtering system can be used to manipulate the input intensities.

The principal element of the variable grating mode device is a thin layer of liquid crystal that is observed to form periodic stripe domains in the presence of an applied voltage. The formation of the domains results in a phase grating characterized by a spatial frequency that depends on the magnitude of the voltage across the liquid crystal layer. The grating period can be optically controlled by placing a two-dimensional photoconductive layer in series with the layer of liquid crystal. The structure of the photoactivated device is shown schematically in Figure 26. The sputter-deposited ZnS photoconductor and the liquid crystal layer are sandwiched between indium tin oxide transparent electrodes deposited on optically flat glass substrates. To operate this device, a dc voltage is impressed across the electrodes. The thin film structure is designed to accept most of the drive voltage when the photoconductor is not illuminated, such that the fraction of the voltage that drops across the liquid crystal layer is below the activation threshold of the VGM effect. Illumination incident on a given area of the photoconductor reduces its impedance, thereby increasing the voltage drop across the liquid crystal layer and driving the liquid crystal into its activated state. Thus, due to the VGM effect, the photoconductor converts an input intensity distribution into a local variation of the phase grating spatial frequency. The variation of optical frequency with voltage across the liquid crystal layer is quite linear for a wide range of VGM liquid crystals, as shown in Figure 27. The fundamental origin of the VGM effect is not well understood, and is the

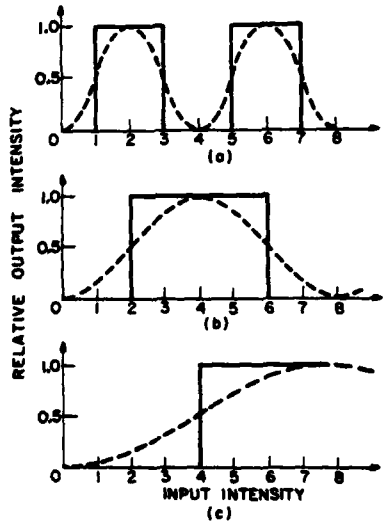


Figure 24. Nonlinear characteristic curves required for the three-bit Gray code. Solid curves are the desired characteristics for the bit plane outputs. Dashed curves are the ideal responses of a multiple period liquid crystal light valve. Parts (a) through (c) represent increasingly significant output bits. (After Armand, Ref. 41).

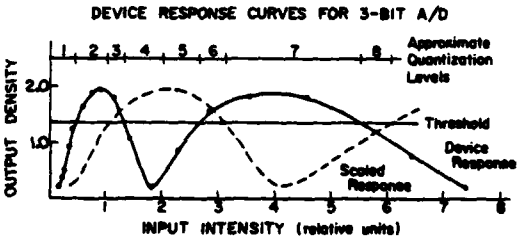


Figure 25. Response curve of the multiple period liquid crystal device used for the three-bit A/D conversion. The solid curve is the measured response. The dashed curve represents the same response with a fixed attenuation of the input. (After Armand, Ref. 41).

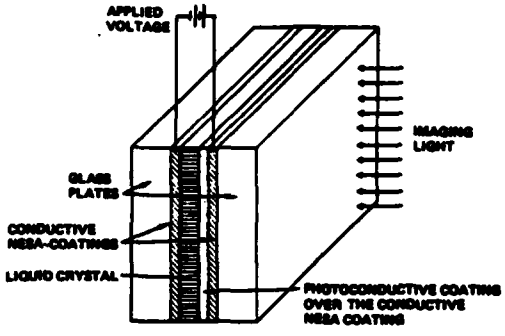


Figure 26. Schematic diagram of the Variable Grating Mode liquid crystal device. (After Soffer, Ref. 43).

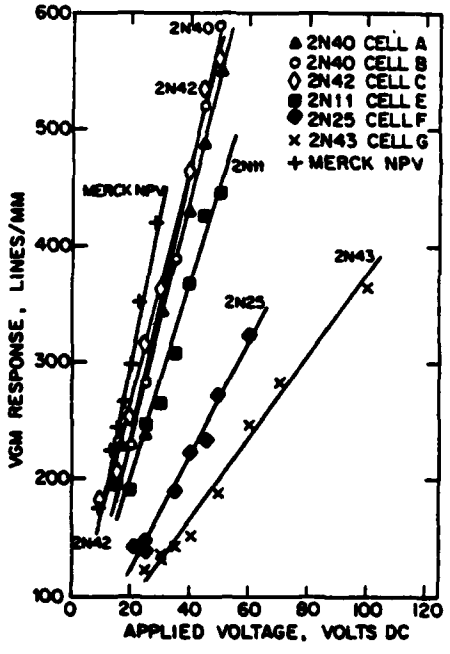


Figure 27. Spatial frequency of VGM domains as a furction of applied voltage for various liquid crystals. (After Soffer, Ref. 43).

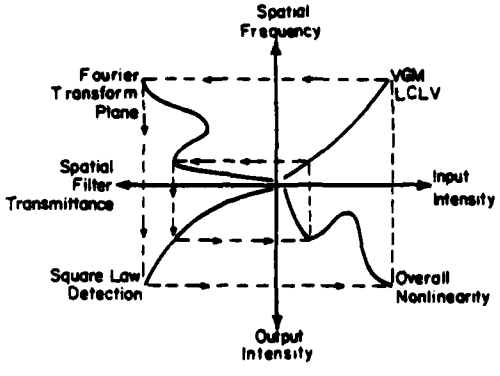


Figure 28. Nonlinear processing utilizing the intensity-to-spatial frequency conversion characteristic of the VGM liquid crystal device. (After Soffer, Ref. 44).

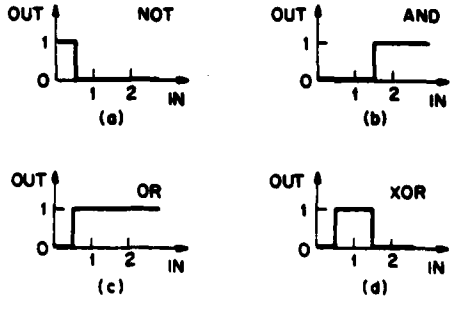


Figure 29. Logic functions as simple nonlinearities. Given an input consisting of the sum of two binary inputs, different logical operations can be effected by means of the depicted nonlinear characteristics. (After Chavel, Ref. 42).

subject of current investigation.

The intensity-to-spatial frequency conversion capability of the VGM device allows the implementation of arbitrary point nonlinearities with simple Fourier plane filters. As discussed above, when an input image illuminates the photoconductive layer of this device, the intensity variations of the input image change the local grating frequency. If conerent light is utilized to Fourier transform the processed image, different spatial frequency components (corresponding to different input intensities) of the encoded image appear at different locations in the Fourier plane. Thus, by placing appropriate spatial filters in this plane it is possible to obtain different transformations of the input intensity in the output plane as depicted in Figure 28. This figure describes the variable grating mode nonlinear processing function graphically. The input intensity variation is converted to a spatial frequency variation by the characteristic function of the VGM device (upper right-hand quadrant). These variations are Fourier transformed by the optical system and the spectrum is modified by a filter in the Fourier plane (upper left-hand quadrant). Finally, a square-law-detection produces the intensity observed in the output plane (lower left-hand quadrant). Considered together, these transformations yield the overall nonlinearity (lower right-hand quadrant). Design of a proper spatial filter for a desired transformation is a relatively easy task. For example, a level slice transformation requires only a simple slit that passes a certain frequency band or bands.

To visualize how the VGM device can be used to implement logic operations, one need only realize that the function of a logic circuit can be represented as a simple binary nonlinearity. The input-output characteristics of the common logic functions are shown in Figure 29. The input in this figure is the simple arithmetic sum of two input image intensities corresponding to logic levels 0 or 1, as shown in the experimental arrangement depicted in Figure 30. For example, NOT is simply a hard-clipping inverter, AND and OR are hard-clippers with different thresholds and XOR is a level slice function. In the VGM approach, such binary nonlinearities can be directly implemented by means of simple slit apertures (41). Thus the particular binary logic function implemented is fully programmable merely by altering the (low resolution) Fourier plane filter. In addition, the input and output functions are physically separate, which provides for the possibility of level restoration of degraded inputs. This feature is essential to the production of a reliable logic system that is immune to noise and systematic errors.

Many logic functions which would normally require multiple gates to implement can be obtained directly with a single VGM cell. An important example is the full-adder where two input bit planes and the carry bit plane are imaged simultaneously onto the VGM device, generating four possible input intensity levels as shown in Figure 31. The four resulting diffracted orders can be filtered to generate the sum bit plane using the positive orders and simultaneously the carry bit plane using the negative orders. Thus, a full addition can be performed in a single pass through the device. Functions requiring matrix-addressable look-up tables can be generated by utilizing two orthogonally oriented VGM devices in conjunction with a two-dimensional Fourier plane filter. Several such functions are required for optical implementation of residue arithmetic. Finally, in addition to the combinatorial logic functions discussed above, sequential logic may also be implemented with appropriate feedback.

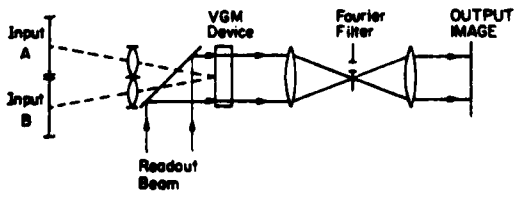


Figure 30. Experimental arrangement for performing logical operations on 2-D binary inputs with a VGM device. (After Chavel, Ref. 42).

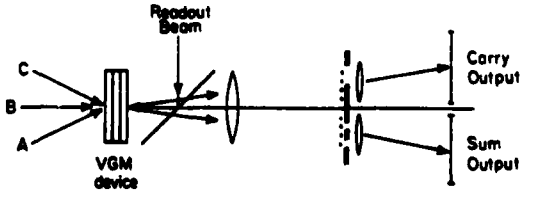


Figure 31. Implementation of a single pass full adder with a VGM device. The inputs A and B represent two binary images (bit-planes) to be added, while input C represents the carry-bit-plane from the previous operation.

The wide range of nonlinear optical processing applications described above demonstrates the tremendous flexibility inherent in the intensity-to-spatial frequency conversion process. It is unfortunate that at present, the VGM device represents the only available real time implementation of this operation. The present major shortcomings of the device are grating imperfections (which give rise to broadening of the diffracted orders), speed of response (on the order of one second), and lifetime. Improvements in all of these areas may be anticipated, since the VGM liquid crystal device is at an early stage of development.

V. Photorefractive Spatial Light Modulators

Until quite recently, the recording of volume holograms in photorefractive materials has been primarily investigated for applications in archival storage and high resolution holographic memories. Applications to coherent optical processing were limited by extremely low writing sensitivities in available electrooptic materials. Recently, however, a number of electrooptic materials have been investigated which exhibit holographic recording sensitivities comparable to that of photographic film, including iron-doped lithium niobate (45), strontium barium niobate (46), bismuth silicon oxide and bismuth germanium oxide (47), and barium titanate (48). The availability of appropriate materials for real time volume holographic storage has spawned interest in several optical data processing applications, incuding phase-conjugate wavefront generation (49), double-exposure and time average holographic interferometry for non-destructive testing (50,51), real time correlation/convolution (52,53), and edge enhancement (54,55). The use of photorefractive materials as spatial light modulators necessitates the use of coherent input and output beams, and as such represents a departure from the traditional function of incoherent-to-coherent conversion in SLM's. However, such devices may be utilized as Fourier plane holographic filters in conjunction with an incoherent-to-coherent or electron-beam-addressed SLM, and in applications (such as multiple exposure holographic interferometry and phase-conjugate wavefront generation) requiring coherent sources.

The physical origin of the photorefractive effect is shown schematically in Figure 32 (56). The intensity interference pattern of two monochromatic coherent plane waves with angular separation 2θ is characterized by a grating vector $k = 4\pi(\sin\theta)/\lambda$ oriented perpendicular to the acute bisector of the plane wave propagation vectors. Within the photorefractive material, either electron-hole pair generation, or electron (hole) excitation from trap states to the conduction (valence) band, or both may occur at a rate proportional to the local intensity. Free carriers so created will diffuse due to the spatially-varying concentration gradient, and will subsequently be trapped preferentially in regions of lower intensity. This charge redistribution can be enhanced by application of an electric field parallel to the grating wave vector. The resultant spatial variation in the charge distribution replicates the grating spacing of the intensity interference pattern, generating a periodic modulation of the local space charge field, which in turn modulates the local refractive indices through the Pockels (electrooptic) effect. The refractive index grating so formed creates a volume phase hologram within the bulk of the electrooptic material, which can be read out by a third monochromatic beam of appropriate polarization counterpropagating along either of the two writing beams. Both transmission and refection holograms may be stored dependent on the orientation of the crystal with respect to the writing beams, as well as on the included angle 2θ. The sensitivity and maximum diffraction efficiency of the photorefractive holographic storage process are functions of the density of carriers available for photoexcitation, the density of available traps for charge redistribution, the magnitude of the electrooptic coefficient, the relative intensities of the writing beams, the magnitude of the applied field (if any), the mobility-lifetime product of the liberated photocarriers, and the grating period, in addition to numerous geometrical factors. Simultaneous optimization of these considerations places many constraints on the selection of appropriate electrooptic materials, and consequently emphasizes continued research on desirable material modifications (primarily doping and improvements in growth techniques for enhanced optical quality) and characterization of the relevant optoelectronic properties of electrooptic materials.

Photorefractive materials have several notable advantages for coherent optical processing applications. Since both image plane and Fourier plane holograms can be recorded will equal ease, great flexibility in optical processing configurations can be made use of. For example, an optical system that is capable of performing real time correlations and convolutions is shown schematically in Figure 33 (52). A similar system configuration that allows implementation of image edge enhancement is shown schematically in Figure 34 (55). In this latter application, use is made of the nonlinear dependence of the diffraction efficiency on the modulation index of the two writing beams by adjusting the reference beam intensity to lie between the bright and dark object intensity levels to enhance the diffraction efficiency in the transition regions where the modulation index approaches unity. A second notable feature of photorefractive materials is extremely high resolution, as depicted graphically in Figure 35. The marked dependence of the amplitude modulation function on the magnitude of the applied field (see Figure 35) allows the possibility of

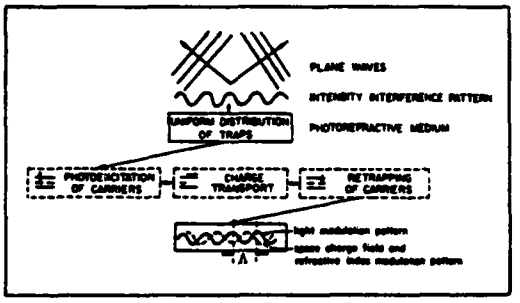


Figure 32. Schematic explanation of the photorefractive effect. (After Kim, Ref. 56).

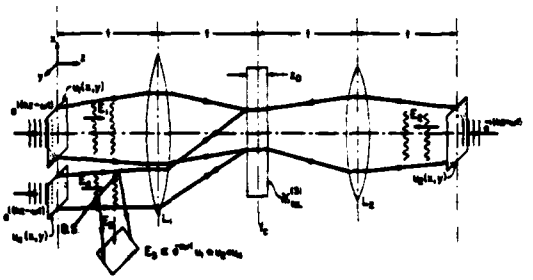


Figure 33. Real time degenerate four-wave mixing convolution/correlation geometry. All input optical fields are at frequency ω. The beam splitter (BS) is necessary to view the desired output, E_3, which is evaluated at a plane located a distance f from lens L_1. (After Pepper, Ref. 52).

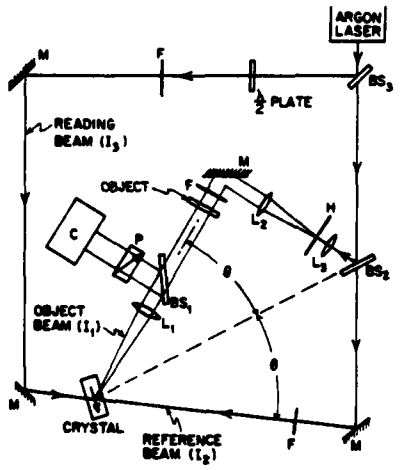


Figure 34. Experimental setup for real time edge enhancement using the photorefractive effect in $BaTiO_3$. Writing beams with intensities I_1 and I_2 (ordinary polarization) and reading beam with intensity I_3 (extraordinary polarization) are shown, as is the C axis of the $BaTiO_3$ crystal. (After Feinberg, Ref. 55).

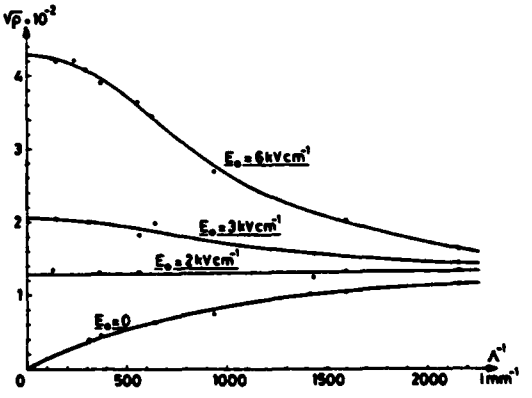


Figure 35. Amplitude modulation transfer function for photorefractive volume holographic storage in $Bi_{12}SiO_{20}$. (After Huignard, Ref. 57).

achieving spatial light modulation with essentially flat MTF characteristics out to spatial frequencies in excess of 1000 line pairs/mm.

VI. Future Directions for Research

In this concluding section, a number of potentially profitable directions for research are presented in addition to those indicated in preceeding sections. Research on the fundamental physical limitations of each candidate spatial light modulator technology is critical both to optimization of current device design and performance, and to meaningful performance comparisons irrespective of current technological limitations. Such comparisons are essential in the assessment of the ultimate potential of each technological approach. This research must be broadly extended to include increased efforts in the materials growth, deposition, processing, and characterization areas, since most of the devices discussed above have been shown to require relatively unique materials properties simultaneously in several distinct materials classes (e.g., the liquid crystal light valves require high resistivity photoconductors, thin film blocking layers, thin film dielectric mirrors, and appropriate liquid crystals). Very little work to date has been reported on attempts to develop high speed spatial light modulators for high frame rate applications. All of the approaches described above are inherently slow and cannot be expected to significantly exceed TV frame rates. Limited frame rate capability will become a major bottleneck in the development of advanced real time optical processing systems and subsystems. Research on the optoelectronic properties of photorefractive materials may yield improvements in both sensitivity and maximum diffraction efficiency. Realization of such improvements would result in widespread availability of inexpensive, real time holographic storage and processing devices. The enormous processing flexibility for linear, nonlinear, combinatorial logic, and sequential logic point operations inherent in the intensity-to-spatial frequency conversion process will hopefully stimulate other possible approaches in addition to that offered by the variable grating mode liquid crystal device. In the area of real time parallel nonlinear optical processing, strong demand exists for high quality two-dimensional variable level slice and threshold functions. Although recent progress in optical bistability has been substantial, most current approaches are either one-dimensional or even single channel. Many "linear" optical processing operations such as correlation, convolution, and Fourier plane filtering require some form of threshold at the output for eventual system implementation.

Acknowledgements

Valuable discussions with C. Warde, D. Casasent, R. Aldrich, J. Grinberg, U. Efron, A.A. Sawchuk, T. Strand, and J. Feinberg are gratefully acknowledged. Research on the physics of operation, materials growth and characterization, and device characterization of real time spatial light modulators is supported in part by the National Science Foundation, the Joint Services Electronics Program, the Air Force Office of Scientific Research, the Army Research Office, and the Ford Motor Company.

REFERENCES

1. D. Casasent, "Spatial Light Modulators," Proc. IEEE, 65, 143-157, (1977).

2. D. Casasent, "Recyclable Input Device and Spatial Filter Materials," in Laser Applications, M. Ross, Ed., Academic Press, New York, (1977).

3. D. Casasent, "Materials and Devices for Coherent Optical Computing," in Optical Information Processing, G.W. Stroke et al. Eds., Plenum Press, New York, (1976).

4. S. Lipson, "Recyclable Incoherent-to-Coherent Image Converters," in Advances in Holography, N. Farhat, Ed., Marcel Dekker, New York, (1976).

5. G. Knight, "Interface Devices and Memory Materials," in Optical Data Processing, S.H. Lee, Ed., Springer-Verlag, Heidelberg, Germany, (1976).

6. J. Flannery, Jr., "Light-Controlled Light Valves," IEEE Trans. Elec. Dev., Ed-20, 941-953, (1973).

7. K. Preston, Jr., Coherent Optical Computers, McGraw Hill, New York, (1972).

8. D. Casasent, "Performance Evaluation of Spatial Light Modulators," Appl. Opt., 18, 2445-2453, (1979).

9. B.A. Horwitz and F.J. Corbett, "The PROM-Theory and Applications for the Pockels Readout Optical Modulator," Opt. Eng., 17, 353-364, (1978).

10. C. Warde, A.D. Fisher, D.M. Cocco, and M.Y. Burmawi, "Microchannel Spatial Light Modulator," Opt. Lett., 3, 196-198, (1978).

11. C. Warde, A. Weiss, and A. Fisher, "$LiTaO_3$ and $LiNbO_3$ Microchannel Spatial Light Modulators," Proc. SPIE Los Angeles Technical Symposium, 218, 59-66, (1980).

12. D. Casasent, "E-Beam DKDP Light Valves," Opt. Eng., 17, 344-352, (1978).

13. G. Marie, J. Donjon, and J.-P. Hazan, "Pockels Effect Imaging Devices and Their Applications," in Advances in Image Pickup and Display. Vol. 1, B. Kazan, Ed., Academic Press, New York, (1974), 225-302.

14. D. Casasent, S. Natu, T. Luu, G. Lebreton, and E. DeBazelaire, "New Birefringence Theory and Uses of the Photo-DKDP Spatial Light Modulator in Optical Data Processing," Proc. SPIE, 202, 122-131, (1980).

15. J. Donjon, F. Dumont, M. Grenot, J.-P. Hazan, G. Marie, and J. Pergrale, "A Pockels-Effect Light Valve: Phototitus. Applications to Optical Image Processing," IEEE Trans. Elec. Dev., ED-20, 1037-1042, (1973).

16. G.J. Berzins and M. Graser, Jr., "Response of a $Bi_{12}SiO_{20}$ Pockels Readout Optical Modulator to X-rays," Appl. Phys. Lett., 34(8), 500-502, (1979).

17. J.C.H. Spence and A. Olsen, "Use of Pockels Readout Optical Modulators (PROMs) for Atomic Resolution Electron Image Processing," Proc. SPIE Los Angeles Technical Symposium, 218, 154-160, (1980).

18. A.R. Tanguay, Jr., "The Czochralski Growth and Optical Properties of Bismuth Silicon Oxide," Thesis, Yale University, (1977).

19. A.R. Tanguay, Jr. and R.C. Barker, "Implications of Concurrent Optical Activity and Electric Field Induced Birefringence for Pockels Readout Optical Memory Performance," 1978 Annual Meeting of the Optical Society of America, San Francisco, California, (1978); to be published.

20. Y. Owechko and A.R. Tanguay, Jr., "Effects of Charge Dynamics and Device Parameters on the Resolution of Electrooptic Spatial Light Modulators," Proc. SPIE, 202, 110-121, (1979).

21. Y. Owechko and A.R. Tanguay, Jr., "Exposure-Induced Charge Distribution Effects on the MTF of Electrooptic Spatial Light Modulators," Proc. SPIE Los Angeles Technical Symposium, 218, 67-80, (1980).

22. Y. Owechko and A.R. Tanguay, Jr., "Theoretical Resolution Limitations of Electrooptic Spatial Light Modulators," 1979 Annual Meeting of the Optical Society of America, Rochester, New York, (1979); to be published.

23. J.C. Dainty and R. Shaw, Image Science: Principles, Analysis, and Evaluation of Photographic-Type Imaging Processes, Academic Press, New York, (1974), 140.

24. R.A. Sprague, "Effect of Bulk Carriers on PROM Sensitivity," J. Appl. Phys., 46 (4), 1673-1678, (1975).

25. C. Warde, Private Communication.

26. D. Casasent and T.K. Luu, "Photo-DKDP Light Valve in Optical Data Processing," App. Opt., 18, 3307-3314, (1979).

27. W.P. Bleha, L.T. Lipton, E. Wiener-Avnear, J. Grinberg, P.G. Reif, D. Casasent, H.D. Brown, and B.V. Markevitch, "Application of the Liquid Crystal Light Valve to Real-Time Optical Data Processing," Opt. Eng., 17(4), 371-384, (1978).

28. P.G. Reif, A.D. Jacobson, W.P. Bleha, and J. Grinberg, "Hybrid Liquid Crystal Light Valve-Image Tube Devices for Optical Data Processing," Proc. SPIE, 83, 34-43, (1976).

29. R.C. Jones, "A New Calculus for the Treatment of Optical Systems. III. The Sohncke Theory of Optical Activity," J. Opt. Soc. Am., 13, 500-503, (1941).

30. G.P. Montgomery, Jr., "Effect of Liquid-Crystal Thickness on the Optical Performance of a Liquid Crystal Image Transducer," Proc. SPIE, 202, 103-109, (1979).

31. G.P. Montgomery, Jr., "Optical Properties of a Liquid Crystal Image Transducer at Normal Incidence, I. Mathematical Analysis and Application to the Off-State," J. Opt. Soc. Am., 70, 287-300, (1980).

32. G.P. Montgomery, Jr., "Optical Properties of a Liquid Crystal Image Transducer at Normal Incidence. II. The On-State," J. Opt. Soc. Am., 70, 843-856, (1980).

33. L.M. Fraas, J. Grinberg, W.P. Bleha, and A.D. Jacobson, "Novel Charge-Storage-Diode Structure For Use With Light-Activated Displays," J. Appl. Phys., 47(2), 576-583, (1976).

34. L.M. Fraas,, W.P. Bleha, J. Grinberg, and A.D. Jacobson, "AC Photoresponse of a Large-Area Imaging CdS/CdTe Heterojunction," J. Appl. Phys., 47(2), 584-590, (1976).

35. S. Natu and D. Casasent, "Sensitometry Control of a Hybrid Field-Effect Liquid Crystal Light Valve," Opt. Comm., 31(2), 135-138, (1979).

36. D. Casasent and S. Natu, "Spatial Variations in a Hybrid Field-Effect Liquid Crystal Light Valve," Appl. Phys., 20, 171-174, (1979).

37. A.D. Gara, "Phase Response of a Liquid Crystal Image Transducer," Appl. Opt., 17(23), 3696-3698, (1978).

38. P.O. Braatz, K. Chow, U. Efron, J. Grinberg, and M.J. Little, "A Fast Silicon Photoconductor-Based Liquid Crystal Light Valve," Proc. Int. Elec. Dev. Mtng., 540-542, (1979).

39. J. Grinberg and U. Efron, Private Communications.

40. J. Grinberg, W.P. Bleha, P.O. Braatz, K. Chow, D.H. Close, A.D. Jacobson, M.J. Little, N. Masseti, R.J. Murphy, J.G. Nash, and M. Waldner, "Liquid-Crystal Electro-Optical Modulators for Optical Processing of Two-Dimensional Data," Proc. SPIE, 128, 253-266, (1977).

41. A. Armand, A.A. Sawchuk, T.C. Strand, D. Boswell, and B.H. Soffer, "Real-Time Parallel Optical Analog-to-Digital Conversion," Opt. Lett., 5(3), 129-131, (1980).

42. P. Chavel, A.A. Sawchuk, T.C. Strand, A.R. Tanguay, Jr., and B.F. Soffer, "Optical Logic With Variable Grating Mode Liquid Crystal Devices," Opt. Lett., 5, 398-400, (1980).

43. B.H. Soffer, D. Boswell, A.M. Lackner, A.R. Tanguay, Jr., T.C. Strand, and A.A. Sawchuk, "Variable Grating Mode Liquid Crystal Device for Optical Processing," Proc. SPIE, 218, 81-87, (1980).

44. B.H. Soffer, D. Boswell, A.M. Lackner, P. Chavel, A.A. Sawchuk, T.C. Strand, and A.R. Tanguay,Jr., "Optical Computing with Variable Grating Mode Liquid Crystal Devices," Proc. SPIE, 232, (1980) (in press).

45. D. Staebler and W. Philips, "Fe-Doped $LiNbO_3$ for Read-Write Applications," Appl. Opt., 13, 788-794, (1974).

46. K. Megumi, H. Kozuka, M. Kobayashi, and Y. Furuhata, "High-Sensitive Holgoraphic Storage in Ce-Doped SBN," Appl. Phys. Lett., 30(12), 631-633, (1977).

47. J.P. Huignard and F. Micheron, "High-Sensitivity Read-Write Volume Holographic Storage in $Bi_{12}SiO_{20}$ and $Bi_{12}GeO_{20}$ Crystals," Appl. Phys. Lett., 29(9), 591-593, (1976).

48. J. Feinberg, D. Heiman, A.R. Tanguay, Jr., and R.W. Hellwarth, "Photorefractive Effects and Light-Induced Charge Migration in Barium Titanate," J. Appl. Phys., 51(3), 1297-1305, (1980).

49. J.P. Huignard, J.P. Herriau, P. Auborg, and E. Spitz, "Phase-Conjugate Wavefront Generation Via Real-Time Holography in $Bi_{12}SiO_{20}$ Crystal," Opt. Lett., 4(1), 21-33, (1979).

50. J.P. Huignard and J.P. Herriau, "Real-Time Double-Exposure Interferometry With $Bi_{12}SiO_{20}$ Crystals in Transverse Electrooptic Configuration," Appl. Opt., 16(7), 1807-1809, (1977).

51. J.P. Huignard, J.P. Herriau, and T. Valentin, "Time Average Holographic Interferometry With Photoconductive Electrooptic $Bi_{12}SiO_{20}$ Crystal," Appl. Opt., 16(11), 2796-2798, (1977).

52. D.M. Pepper, J. AuYeung, D. Fekete, and A. Yariv, "Spatial Convolution and Correlation of Optical Fields Via Degenerate Four-Wave Mixing," Opt. Lett., 3(1), 7-9, (1978).

53. J.O. White and A. Yariv, "Real-Time Image Processing Via Four-Wave Mixing in a Photorefractive Medium," Appl. Phys. Lett., 37(1), 5-7, (1980).

54. J.P. Huignard and J.P. Herriau, "Real-Time Coherent Object Edge Enhancement with $Bi_{12}SiO_{20}$ Crystals," Appl. Opt., 17(17), 2671-2672, (1978).

55. J. Feinberg, "Real-Time Edge Enhancement Using the Photorefractive Effect," Opt. Lett., 5(8), 330-332, (1980).

56. D.M. Kim, T.A. Rabson, R.R. Shah, and F.K. Tittel, "Photorefractive Materials for Optical Storage and Display," Opt. Eng., 16(2), 189-196, (1977).

57. J.P. Huignard, J.P. Herriau, G. Rivet, and P. Gunter, "Phase-Conjugation and Spatial-Frequency Dependence of Wave-Front Reflectivity in $Bi_{12}SiO_{20}$ Crystals," Opt. Lett., 5(3), 102-104, (1980).

Discussion (Armand R. Tanquay, Jr.; Discussion Leader: Cardinal Warde)

Q. You said the Hughes light valve might have a resolution of 150 lines/mm. How large can such a device be fabricated?

A. 1" x 1".

Q. Is that a physical limitation?

A. No, but with 150 lines/mm resolution you don't need more than that. It is a technological limitation in that you need 5 mil thin, high resistivity wafers that you have to get uniform to within a percent. Holding onto a 5 mil silicon wafer is a trick in itself. With a 1" x 1" wafer, you can use a blocking wafer to give it support, but with a 2" x 2" wafer you have serious problems. However, if you are going to buy 4 or 5 million of them, I'm sure the people in silicon technology would come through.

Q. Do you have any data on the repeatability of the variable grating mode experiments? Is temperature sensitivity a problem?

A. Temperature dependence is very tiny. The best experiment I can give you there is by putting the device under an incandescent lamp and a polarizing microscope for 4 or 5 hours, you don't see any smearing or deformities. I want to be very careful to state the case on the variable grating mode device. This is an exciting device, mostly because of the intensity to spatial frequency conversion that is being done. It is frightening to us working on it to think about the number of problems that need to be overcome. On the other hand, in the liquid crystal business, the situation with digital watches a few years ago was that they had a 4 or 5 second turn on time, until someone finally spent the money to sit down and determine the nature of the situation. When they succeeded in this, the time came down from several seconds to on the order of microseconds. We don't know what the nature of the variable grating mode is at all. We know a lot of things about it, but we don't know what the origin is.

White light optical processing

E. N. Leith, J. Roth and G. Swanson

Department of Electrical and Computer Engineering
The University of Michigan, Ann Arbor, Michigan 48109

Abstract

Optical systems for performing holography and optical processing are described. These systems are coherent, but operate with temporally incoherent light. The noise suppression mechanism of the system is analyzed. In particular, a method for deblurring an image is given.

Introduction

Optical processing systems generally fall into two categories: coherent and incoherent. Our work is on a third ground, which partakes of both. We use temporally incoherent light in an achromatic coherent optical system.

As is well known, coherent light has significant advantages. The use of coherent light enables optical information processing systems to perform sophisticated operations, such as spatial matched filtering. However, the use of coherent light results in the introduction of so-called coherent noise. This noise, the manifestation of light from scatterers outside the plane of the object, is unavoidable, but can be minimized by the use of clean techniques. Even so, images formed in coherent light are always marred in appearance by this noise.

This problem can be reduced by the use of incoherent techniques. For example, there exist methods for making holograms and spatial filters with incoherent light. These techniques work in accordance with the principles of incoherent optics but, in general, such incoherent systems have problems even more severe than their coherent analogues. There is the so-called bias-buildup problem, where, as the number of object points is increased, the contrast of the resultant interference fringes decreases. Since these fringes constitute the recorded signal, the signal amplitude is greatly decreased, with the result that again the signal-to-noise ratio is low.

In our work, we carry out information processing using temporally incoherent light, but in a manner such that the light obeys the principles of coherent rather than incoherent, optics. That is, the optical system is linear in complex amplitude rather than intensity. In this manner we utilize the enromous advantages of coherent processing, without incurring the associated noise problems.

The need for temporal coherence in holography and optical processing can be described very simply: When a polychromatic source is used in forming an interference (or diffraction) pattern, such as in holography, the various wavelength components produce fringe patterns which do not coincide; hence the resultant fringe pattern, the superposition of the fringes formed by all wavelengths, will generally be reduced in contrast. The task in eliminating the temporal coherence requirement, then, is to modify the fringe-forming system so that the fringes formed by each wavelength component are scaled alike and are in registry.

We describe an optical system which accomplishes this achromatization. This system can be implemented either as an optical data processing system or as a hologram recording system. Indeed, as will be seen, in their implementation the distinction between the holographic system and the image processing system disappears.

The basic method

The basic system for performing either holography or optical processing is shown in Figs. 1 and 2. Figure 1 applies to the holographic application. The transparency $f(x)$ is illuminated with a polychromatic light source that is wavelength-dispersed along the longitudinal, or z, axis. For wavelength λ_1, the source is a point a distance z_{s1} from the object; for a wavelength λ_2, the source is a different distance, z_{s2}, from the object. The source may be written as the function

$$S = \delta(z - z_s\lambda_0/\lambda), \tag{1}$$

which indicates that the source point, as a function of wavelength, is located at $z=(\lambda_0/\lambda)z_s$, where z_s is the distance between source and object for wavelength λ_0. Such a source can be

produced by illuminating a Fresnel zone plate with a point white light source and selecting one order while removing the others by appropriate spatial filtering.

We proceed to show how to form a hologram with this arrangement. For simplicity, we form first a simple Gabor in-line hologram. Illuminating the object transparency is a beam $\exp(i\pi x^2/\lambda z)$, which, using the relation $z=(\lambda_0/\lambda)z_s$, becomes $\exp(i\pi x^2/\lambda_0 z_s)$; thus, the form of the illuminating function is now wavelength independent. A lens L forms the Fourier transform of the product $[f(x)\exp(i\pi x^2/\lambda_0 z_s)]$, and we observe in the back focal plane the field distribution

$$f'(x') = \mathcal{F}[f(x - x')\exp(i\pi x^2/\lambda_0 z_s)], \tag{2}$$

where x' describes movement of f(x) through the optical system aperture. However, we confine our observation to a point on axis by placing at the back focal plane of the lens a mask containing a pinhole. The field in the pinhole represents the integral

$$f'(x') = \int f(x - x')\exp(i\pi x^2/\lambda_0 z_s)dx = f*h. \tag{3}$$

We have thus formed the convolution operation describing Fresnel diffraction, and it is significant to note that it has been formed achromatically, using a broad-spectrum source. Of course, it is generated one point at a time, and to generate the entire pattern, we must move the object transparency through the aperture.

This diffraction pattern can be recorded as a hologram by introducing a reference beam $u_0=a_0\exp(i2\pi f_0 x')$, producing an irradiance $|u_0+f'|^2$, which results in a hologram that produces the desired reconstructed wave. By using an achromatic interferometer to provide the two beams, object and reference, the recombination of the beams is accomplished achromatically, as we explain later.

Explicitly showing the zone plate which provides the dispersed source for the object leads to an alternative way (Fig. 2) of describing this holographic process. The zone plate, whose amplitude transmittance is $t_a=1/2[1+\cos(\pi x^2/\lambda_0 z_s)]$, is imaged onto the transparency f(x). We suppose that in the process the zone plate is spatially filtered so that only the virtual image term reaches the transparency f(x). Thus, the light distribution impinging on the transparency is, to within a constant, $\exp(i\pi x^2/\lambda_0 z_s)$, which is the result

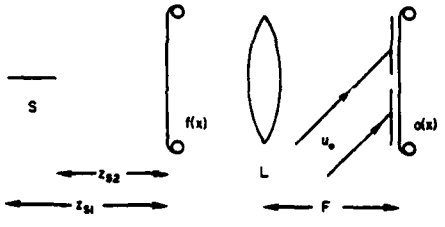


Fig. 1. Achromatic holographic configuration.

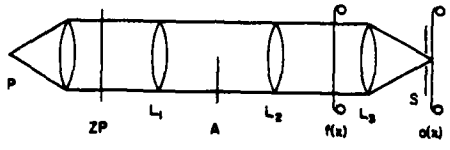


Fig. 2. Alternate holographic configuration.

given previously. A more complete description is given in Refs. 1 and 2. By formulating the operation in this manner, it is seen to be just the basic method of using a coherent optical system to convolve two functions by imaging one onto the other. The holographic process is then just the special case of choosing the convolving impulse response as a quadratic phase function. We could use any other convolving function; thus, the method we describe in terms of holography is indeed a very general method for using optical processing for performing the general operation g=f*h, where f and h are any functions.

Noise-suppression characteristics

There are a number of noise-suppression mechanisms inherent in the achromatic system. The most obvious and simple one is that since we take the output at a single point, spatial [illegible] at any arbitrary plane in the optical system will be transformed into a spatial [illegible] at the output and will become just a constant n'(0,0) at the output point. [illegible] mechanism does not depend on the broadband illumination and would be the [illegible] optical correlator which used point readout instead of spatial readout. It [illegible] presumably offer noise-suppression performance as a hologram-forming system [illegible] the usual holographic systems, which utilize spatial readout. The noise

problem, however, is by no means eliminated by the point readout, and perhaps not even significantly reduced, for several reasons. First, in some cases in optical processing, the output in such "point output" correlators is not taken from a single point, but from some surrounding region, for which the achromatic and other requirements are approximately correct. Second, in some cases the operation to be performed between object and output planes is one-dimensional, with an imaging process being performed in the other dimensions. Thus, the output is taken along a line, and there will in general be a noise spatial fluctuation n'(0,y) [or n'(x,0)] along this line. Finally, much of the noise is signal-dependent noise, resulting from light which interacts with both the object transparency and the noise scatterers, and will thus vary as the output film record is scanned past the output point.

In general, the noise suppression of the achromatic coherent system is dependent upon the specific operation that the system performs. We now begin discussion of the noise-suppression mechanisms for the achromatic system for holography.

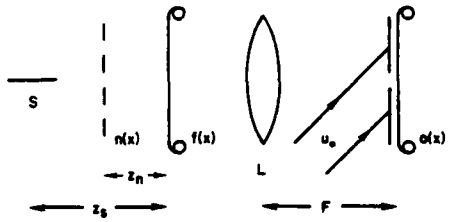


Fig. 3. Configuration for noise analysis.

The signal f(x), which is to be recorded as a hologram, will in practice contain a bias term f_b and a spatially varying term $f_s(x)$. Thus, f(x) is of the form $f(x)=f_b+f_s(x)$. As the light from the noise scatterers passes through the signal plane, it interacts with the bias term to continue as signal-independent noise and also produces signal-dependent noise through interaction with the spatially varying term.

Now consider the reduction of the signal-independent noise achieved by broadening the spectral bandwidth of the source. Let the noise be any variation of transmittance in a plane other than the object and output planes. Such noise could, for example, be from dust, scratches on glass, bubbles in glass, and so forth. Considering the achromatic system of Fig. 1, let the noise originate in a plane a distance z_n from the object plane, as shown in Fig. 3. The field illuminating this plane is, for any wavelength λ,

$$u_n^-(x) = \exp[i\pi x^2/\lambda(z + z_n)], \tag{4}$$

which can be written

$$u_n^-(x) = \exp[i\pi x^2/(\lambda_0 z_s + \lambda z_n)]. \tag{5}$$

Consider for the moment, for purposes of analysis, this noise source to move through the optical system, as does the object signal. Then the field just to the right of the noise plane is

$$u_n^+(x) = n(x - x')\exp[i\pi x^2/(\lambda_0 z_s + \lambda z_n)], \tag{6}$$

where the x' coordinate is the shift of the object signal (and the output signal) with respect to the optical system. The output amplitude, for any wavelength λ, taken at the achromatic point in the output plane becomes

$$u_{out}(x',\lambda) = \int n(x - x')\exp[i\pi x^2/(\lambda_0 z_s + \lambda z_n)]dx, \tag{7}$$

where here the wavelength dependence is shown explicitly to emphasize that the output amplitude varies as a function of λ. Note that the output does not depend on the distance between the signal plane and the correlating lens.

Using again the Fourier transform relation $n*h = \mathcal{F}\{NH\}$, we can rewrite Eq. (7) as

$$u_{out}(x',\lambda) = N(f_x) \times \exp[-i\pi(\lambda_0 z_s + \lambda z_n) f_x^2] \exp(i2\pi f_x s') df_x. \quad (8)$$

In the formation of a hologram, this noise signal, along with the object signal, is combined with a reference function, recorded, and then regenerated by the hologram.

If we broaden the source spectrum, integrating Eq. (8) over a band of wavelengths gives

$$\bar{u}_{out}(x',\Delta\lambda) = \int_{\lambda_0-\Delta\lambda/\lambda 2}^{\lambda_0+\Delta\lambda/2} d\lambda \int_{-\infty}^{\infty} df_x N(f_x) \times \exp[-i\pi(\lambda_0 z_s + \lambda z_n) f_x^2] \exp(i2\pi f_x x'), \quad (9)$$

the average output amplitude. Interchanging the order of integration and performing the integration on λ, we find

$$\bar{u}_{out}(x',\Delta\lambda) = \int_{-\infty}^{\infty} N(f_x) H_2(f_x) \exp(i2\pi f_x x') df_x = \mathcal{F}^{-1}[N(f_x) H_2(f_x)], \quad (10)$$

where

$$H(f_x) = \Delta\lambda \exp[-i\pi(\lambda_0 z_s) f_x^2] \operatorname{sinc}[z_n f_x^2 \Delta\lambda/2].$$

We see that this transfer function is centered at the plane $z_n=0$, the plane of the object signal, and decreases for values of z_n outside this plane. Thus under broad source illumination, we suppress noise from all planes, not just planes to one side of the object plane. Also, the signal is not at all suppressed by the broadband spectrum. We have therefore achieved our objective; we have made a coherent optical processing system that behaves like an incoherently illuminated imaging system in its noise suppression.

This system is in sharp contrast to the conventional holographic system, where H is centered at the recording plane under other polychromatic illumination, and signal from the object plane is attenuated.

Introduction of the reference beam

To complete the achromatic holographic system of Fig. 1 (or 2) we must bring in a white light reference beam that is coherent with the signal beam. This is not a simple task. It can be done by achromatic interferometry. The resulting system is shown in Fig. 4. Three diffraction gratings have been incorporated into the system for generation of the Fresnel diffraction pattern. The system is a three grating interferometer, as well as an achromatic optical processor, with both aspects being thoroughly integrated. The theory of the three grating interferometer has been given elsewhere.[3]

In the system of Fig. 4 (a system for making one-dimensionally dispersed holograms, such as multiplex or rainbow holograms), we have an optical processing system identical with that of Fig. 2, except for the addition of a cylindrical lens L_3 (so as to produce

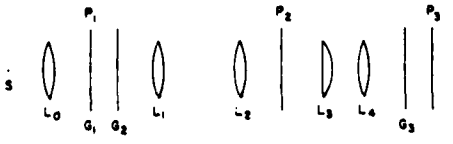


Fig. 4. Three-grating device.
Note that L_3 is a cylindrical lens.

one-dimensional dispersion). G_2 and G_3 are identical gratings of spatial frequency f_1, and G_1 is an off-axis zone plate structure, with carrier f in the y direction and the zone plate structure in the x direction. The zero and first orders of G_1 are selected, with the first order being modulated by the zone plate. G_2 demodulates the diffracted beam to zero spatial frequency in the y direction. Since G_2 affects the light distribution in the y direction only, G_1 can be considered to be imaged in the x dimension at object plane P_2 producing the required object distribution. The input is then processed as described earlier, while G_3 modulates the reference beam in order to produce the required fringes at the output plane. The advantages of this system are:

(a) s(x,y) is imaged in the y dimension at the output plane.

(b) The object beam is undispersed between the object and recording planes. Previous analyses of achromatic grating interferometers for holography indicate that an undispersed object beam is best.

(c) High spatial frequency fringes are possible.

(d) Reference and object beams are spatially separated at the input plane.

Disadvantages of the system are:

(a) The required spatial frequency of the gratings is closely tied to the system magnification. For example, if G_1 and G_3 are of the same spatial frequency, the lens system must produce, to great exactness, unity magnification.

(b) Alignment is very error sensitive.

(c) Light throughput is reduced because of the multiplicity of diffracting elements.

We note that r can be other than a zone plate, which would result in generalized holograms, in which each object point on the hologram becomes a spread function that is other than a zone plate. Also, if using modern techniques for production of phase transparencies we could generate the transparency r_d directly, r_d could be superimposed on s, and the system would thereby be simplified.

Fig. 5. Achromatic fringes (150 lines/mm) produced by the three-grating system

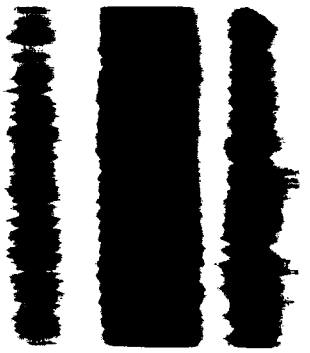


Fig. 6. Output from a slit signal.

Fringes formed by the system, for the case of no signal, are shown in Fig. 5. The spatial frequency of the fringes is in excess of that permitted by the optical transfer function, and the fringes are of high contrast. Fig. 6 shows the diffraction patterns of a slit, along with the fringes formed by the reference beam. This distribution, if recorded, would be the hologram of a slit.

A deblurring system

Finally, we show how to use the achromatic device to deblur images. Here, the system functions as an optical processor only, and the reference beam is not required. This is of course an enormously simpler system than for the case of making a hologram.

The system is shown in Fig. 7. The theory of operation is as follows. We assume the images to be degraded by one-dimensional blurring, such as would arise if the image moved linearly relative to the recording device during the recording process. Let the point-spread function of the blurring process be the rectangular function rect x/L, where L is the width of the blur (here taken along the direction x).

The blurred image, with transmittance s(x,y), can be deblurred to produce an image approximating an original unblurred image by convolving with a restoring function m(x,y),

$$u(x,y) = \int s(\alpha - x,y)\,m(\alpha,y)\,d\alpha \qquad (11)$$

$$= s * m,$$

where α is a dummy variable.

This basic deblurring problem, in which the blur process is characterized by a rect function, has been treated many times. Swindell[4] describes heuristically an algorithm for the restoration process that assumes the function m to be a sequence of pulses of opposite polarity, as shown in Fig. 1. The pulse periodicity is chosen to be exactly the blur width. The convolution is readily carried out mentally by imagining the blurred image of a point (the rectangle s) sliding across m. The convolution integral u is generally zero, with three exceptions. At x=0, the mask is positioned as the center of the pattern, so the blurred image falls on two positive pulses and the integral is therefore not zero. Also, for a displacement x that places the blurred image at the edge of the function m so the blurred image overlaps only the final pulse, the integral is again nonzero. Each response represents the restored image. That each such blurred object point results in three restored images causes no problem provided that the mask m is larger than the extent of the image. Then the various deblurred images will not overlap.

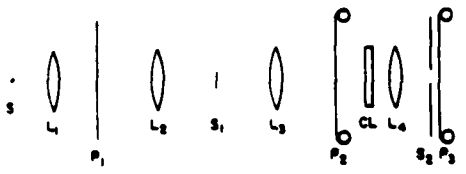


Fig. 7. Optical processing system. S is the light source, L_1; L_2, L_3, and L_4 are lenses; CL is a cylindrical lens; S_1 is a spatial filter (a stop); and S_2 is a slit.

We describe a simple method for carrying out the deblurring process optically, wherein the deblurring function m is produced from simple, readily available diffraction gratings, such as Ronchi rulings.[5] The optical-processing system is shown in Fig. 7. The function m (at P_1) is imaged at s (at P_2), which is moved through the optical-system aperture. Following s is a cylindrical-spherical-lens combination (CL, L_4), followed by a slit in the back focal plane of the second lens. A moving film or other detector records the light in the slit. The system is illuminated with coherent light. This optical system generates the function

$$u(x,y) = s * m, \qquad (12)$$

where the convolution is on x only. The system, in the y dimension, is a straightforward imaging system, with L_3 and L_4 imaging s onto the slit.

The function m is derived from a Ronchi ruling r at the back focal plane of L_1. The Ronchi ruling consists of opaque and transparent bars with sharp transitions [Fig. 9(a)]. The required function m(x) is derived from the grating by a spatial filter that forms the

second derivative of r. We describe (again heuristically) the spatial-filtering process. The first derivative of a square wave [Fig. 9(b)] is ideally a sequence of δ functions located at the transition regions of m and alternating in sign. In Fig. 9(c) we show the second derviative, wherein each pulse in b is converted into two abutting pulses of opposite polarity. In practice, of course, the spatial-filtering process will have a finite spatial-frequency bandwidth.

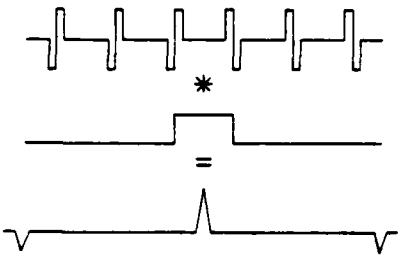


Fig. 8. Deconvolution of a linear blur (after Swindell).

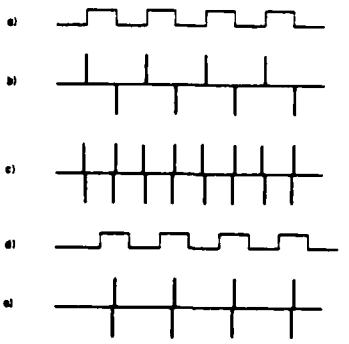


Fig. 9. Derivation of deblurring function from Ronchi grating.

The spatial filter that produces the second derivative is

$$H = Cf_x^2, \tag{13}$$

where C is a constant. Such a mask, appropriately truncated to avoid infinities, is required at the back focal plane of L_1. As a practical matter, a binary approximation to H is satisfactory and is even easier to implement. Such a filter would consist of a stop at the center of the grating diffraction pattern, blocking out some of the lower orders while the higher orders are passed without attenuation.

Closer inspection reveals that the pulse sequence of Fig. 9(c) is not quite right. Alternate pulse pairs are oppositely polarized; one pulse pair, produced from a left-hand edge of the transparent grating slit, has the positive pulse on the left and negative one on the right. The adjacent pulse pairs are the reverse. This problem is corrected by placing a second Ronchi ruling at the signal plane, positioned to block alternate pulse pairs. Thus the signal s sees only the desired function, shown in Fig. 9(e). Correlation with such a function will produce a pair of deblurred images separated by a distance equal to the width of the grating.

This system has been experimentally verified by using a slit to represent the blurred image of a line object. The results are just as theory predicts. We presently are working on a full deblurring experiment.

Thus, we have shown how to perform holography or optical processing using white light, and we have demonstrated a noise reduction mechanism for this process.

Acknowledgements

This work was sponsored by the Air Force Office of Scientific Research (Contract AFOSR 77-3356).

References

1. E. Leith and J. Roth, "White Light Optical Processing and Holography," Appl. Opt., Vol. IC, p. 2565 (1977).
2. E. Leith and J. Roth, "Noise Performance of an Achromatic Coherent Optical System," Appl. Opt., Vol. 18, p. 2803 (1979).
3. E. Leith and G. Swanson, "Achromatic Interferometers for White Light Optical Processing and Holography," Appl. Opt., Vol. 19, p. 638 (1980).

4. W. Swindell, "A Noncoherent Optical Analog Image Processor," Appl. Opt., Vol. 9, p. 2459 (1970).
5. E. Leith, "Image Deblurring Using Diffraction Gratings," Opt. Lett., Vol. 5, p. 70, (1980).

Discussion (Emmett N. Leith; Discussion Leader: Steven K. Case)

Q. Can you tell us something about the diffraction gratings? Are they holographic?

A. Yes. It is all we can reasonably afford, and they are quite suitable because you can tailor them to your needs. You can put however much light you want in any given order and you don't need extremely high quality gratings. However, making them is rather difficult because if you look at the parameters that describe what you need for achromatization, it is very severe. For example, of the three gratings, it turns out that two must be of exactly the same spatial frequency. But that is assuming that the two lenses that are between have exactly the same focal length, and tolerances of lens manufacturers are not that great. So if the lenses differ, you have to make the gratings differ by the same amount. What you do is make a first grating, then to make the second grating you use Moiré techniques to see exactly the spatial frequency that you want. Then you have to take into account magnification by the lenses in the system, because the final grating has to match the spatial frequency of the first grating per se, so the whole thing is fraught with problems.

Real-Time Processing with Acousto-Optical Devices

John N. Lee, N. J. Berg, M. W. Casseday and I. J. Abramovitz

U.S. Army Electronics Research and Development Command, Harry Diamond Laboratories
2800 Powder Mill Road, Adelphi, Maryland 20783

I. Introduction

Light diffracted by an acoustic wave signal will contain the phase, amplitude, and frequency information of that signal, so it would appear that this interaction should be a naturally attractive basis for signal processing. The history of acousto-optic processing could be considered to originate as far back as 1932 with the experimental verification of light diffraction by sound waves by Debye and Sears, since the acousto-optic deflector is the elementary building block for such processing. The practicality of acousto-optic processing was, however, greatly enhanced by a number of technological innovations, especially the advent of the laser, progress in the growth of good crystals of acousto-optic materials, and progress in the fabrication and design of acoustic delay lines. Several of these innovations, aside from their relevance to acousto-optics, have spurred work in information processing along well-known courses, e.g., optical processing via two-dimensional transforms using lenses and spatial light modulations, and signal processing using acoustic filters, especially surface acoustic wave(SAW) filters. It is the purpose of this paper to review acousto-optic approaches to real-time signal processing.

One can cite several general advantages of an acousto-optic approach to real-time signal processing. First, the acousto-optic interaction is approximately linear over a very large dynamic range. Unlike tapped delay-line filters, there is no loading of the delay-line material that could result in acoustic velocity dispersion and signal attenuation. Second, acoustic delay lines can be fabricated with very wide bandwidths (>500 MHz) and long interaction windows (~80 μs). Flexibility in the design of transducers, particularly in the case of SAW's, allows for prefiltering of input signals. Finally, if one uses SAW delay lines, one has a planar technology that is inherently compatible with integrated optics; hence, the optical portion of a processor could also be made compact and rugged.

In the following, a review of the current status of acousto-optic deflectors and their application to spectral analysis will be given, followed by a review of a variety of acousto-optic signal-processing concepts that have been implemented, such as correlators, Fourier transformer devices, and triple-product processors.

II. Acousto-Optic Deflectors and Spectral Analysis

The earliest applications of acousto-optic cells were to the deflection and modulation of light. The parameters for optimum operation of deflectors and modulators are generally at variance:[1] that is, a good modulator will perform poorly as a deflector and vice versa. Since the topic of spectral analysis is a major concern, only acousto-optic deflectors will be considered.

The operation of an acousto-optic deflector depends on the change in deflection angle of the light with change in frequency. Operation is generally in the Bragg regime, and the angle of deflection is therefore governed by the relation

$$\theta = 2 \sin^{-1}\left(\frac{\lambda f}{2nv_s}\right), \qquad (1)$$

where λ is the light wavelength, n the index of refraction of the medium, v_s the acoustic-wave velocity, and f the frequency of the acoustic wave. A major attraction is that all frequencies contained in the acoustic wave are displayed simultaneously. The number of frequencies that may be resolved is limited by diffraction of the light. If $\Delta\theta$ is the angular spread due to a signal of bandwidth Δf, the number of resolvable frequency cells is obtained by dividing $\Delta\theta$ by the diffraction-limited angular spread of the light-beam $\Delta\phi$ ($=\lambda/d$, where d = the width of the light beam). The result is the time-bandwidth (T-B) product, where the relevant time is the transit time of the acoustic signal for a distance d. Other definitions of time-bandwidth product will be given later. With present-day technology, time-bandwidth products of about 2000 to 5000 are possible.[2] Three parameters limit the ultimate T-B product: length of the delay line, attenuation of the acoustic wave by the delay-line material, and diffraction of the acoustic wave. With improved crystal growth techniques, delay lines up to 15 cm long have been grown, and for the 00λ SAW cut of BGO, this corresponds to an aperture of greater than 80 μs.[3] Increasing the aperture by using materials with very slow acoustic velocities is often hampered because the attenuation of the acoustic signal is larger for such materials. For example, TeO_2 has a shear mode velocity of only 620 m/s but unfortunately a large acoustic attenuation (17.9 dB/s at 1 GHz) that limits effective apertures to about 50 μs.[4] Attenuation also increases with the square of the acoustic frequency, thereby limiting bandwidth.

Acoustic-wave diffraction also results in a limit on the bandwidth. Normally the most effective mode of operation for a deflector is to have a planar optical wave incident on the acoustic cell[1] at the Bragg angle. For a single acoustic transducer with a limiting aperture L, the diffraction spread is approximately λ/L, according to scalar diffraction theory, where λ is the acoustic wavelength. By differentiation of eq. (1), one sees that the diffraction spread is equivalent to a bandwidth

$$\Delta f/f_o = 2/\lambda_o^2/\lambda L \qquad (2)$$

where λ_o is the wavelength at center frequency f_o. The bandwidth is therefore a function of the center frequency. $\Delta f/f_o$ can be increased by decreasing L, but only at the expense of interaction efficiency with the optical beam. More desirable is to use phased or tilted arrays of transducers to increase Δf while maintaining L.[5] This technique is especially suited to SAW implementation, and 600-MHz bandwidths have been achieved.[6] Bulk-wave devices have the advantage of higher center frequencies of operation; $LiNbO_3$ devices having 1-GHz bandwidths and 2.3-GHz center frequencies have been demonstrated.[2]

One can use an acousto-optic deflector for rf spectral analysis by placing the acoustic cell at the front focal plane of a lens. The rf power spectrum will be displayed as a light-intensity distribution at the back focal plane of the lens. Photodetector arrays may be placed at the back focal plane for signal detection. The attractiveness of this implementation is nearly instantaneous coverage of the full bandwidth of the device input. However, several considerations should be kept in mind when one employs this apparently simple concept. In any acoustic device there is a finite loading time, so that the ultimate resolution and signal strength will not be reached until the signal has entirely filled the aperture of the acoustic cell. For pulsed input signals, it is possible that the aperture will not be filled, so even when the signal has fully entered the cell, the resultant $(\text{sinc})^2$ light-intensity functions at the detector array (for uniform illumination) will be broader and resolution therefore poorer. Also, if photodetection is done with an array that integrates the light intensity for some fixed amount of time before readout (for processing gain; see section III), the effect of the moving acoustic signal will be to broaden the main peak and smooth the sidelobes of the sinc function.[7]

Perhaps the most serious disadvantage of the spectrum analyzer described above is that although the acousto-optic interaction gives nearly 100-percent time coverage of a frequency band, most detector arrays must be read out serially, and present-day shift registers cannot operate much faster than about 5 MHz. Hence, a fraction of a millisecond must be expended for each readout cycle of a 1000-element array. An obvious solution would be to employ GaAs technology to produce faster readout circuits; a 500-MHz CCD array has been reported.[8] A second approach is to use parallel-processing electronics. A parallel-readout fiber-optic array, with each fiber connected to a discrete detector, is being developed;[9] 288 fiber ends will be placed contiguously, and with the PIN diodes and associated video amplification, a dynamic range of >40 dB and a response time of <1 μs are expected.

If several acoustic signals of differing frequencies are present simultaneously, cross-modulation will result in the generation of intermodulation beams, since the generation of the various diffracted beams will not be completely independent.[10] The case for two acoustic signals has been evaluated; in the Bragg regime it was found that third-order intermodulation reduced the spurious-free dynamic range by 20 dB from a potential 70 dB in the single-signal case to about 50 dB.[10]

The light at the back focal plane of the spectrum-analyzer lens can also be further processed. Both phase and amplitude information are preserved at the back focal plane, since the lens performs a true Fourier transformation on the light beam. Coherent detection at the back focal plane, using a reference light beam, can be used to recreate the time-domain signal inserted into the acoustic cell. Operations such as filtering, weighting, or excision of interfering signals may be performed at the frequency plan before the reconstruction of the signal.

III. Acousto-Optic Processors and Architectures

This section will deal with the various acousto-optic devices other than the simple Bragg-cell spectrum analyzer that have been demonstrated and the performances that have been achieved by them. These devices are mainly correlator/convolvers or systems requiring the computation of the correlation/convolution integral. For convenience, devices will be categorized as space integrating or time-integrating.

Space-Integrating Processors

Correlation or convolution between two signals can be obtained by imaging the diffracted light from one signal onto the other, provided one of the signals is moving with respect to the other. The imaging process, resulting in doubly diffracted light, provides the necessary multiplicative operations. The doubly diffracted light must then be collected and focussed onto a photodetector to effect the integral. This scheme is therefore termed space-integrating; the correlation/convolution function appears at the output of the photodetector as a function of time. Movement of one signal with respect to the other may be achieved by transduction of one or both signals into acoustic signals. The earliest correlators used one acoustic signal and one fixed reference mask.[11] Use of two acoustic signals eliminates the need for fabricating a photographic mask or using a spatial light modulator. However, relative motion between the two signals must be obtained either by propagating the signals in opposing directions or by having different acoustic velocities (i.e., materials) for the two signals. In the former case, a convolution in a compressed time frame is obtained, unless one of the signals is time reversed so as to obtain correlation. In the latter case, correlation is obtained if the length and frequencies of the acoustic signals are scaled inversely with the ratio of the two acoustic velocities.[12] Systems using counterpropagating signals have been constructed with bulk-wave delay lines[13] and with SAW delay lines.[3] A "two-crystal" correlator has been constructed with SAW lines.[12] Figure 1 illustrates the SAW acousto-optic convolver.

Correlators and convolvers may be characterized by their time-bandwidth product, as defined in Section II. It can be shown that the T-B product is essentially proportional to the signal-to-noise processing gain obtainable in correlation,[14] provided the system has dynamic range in excess of this processing gain. It is usually desirable to make this product as large as possible. By using large aperture delay lines and the tilted SAW transducer technique described in section II, the convolver shown in Fig. 1 achieved a T-B product of 10,000 (250-MHz bandwidth, 40-μs aperture), while the two-crystal correlator achieved a T-B product of 3,000.[12] Limitations on T-B product are exactly those discussed in section II. Large dynamic range (>60 dB) is possible with these acousto-optic processors, because the frequency shift of the light upon diffraction allows one to use coherent, heterodyne detection. Hence, in Fig. 1, the light beam A_0, which is undiffracted light, serves as the local oscillator for detecting the amplitude of the doubly diffracted light beam A_2.

A second type of space-integrating correlator which uses memory storage has been developed.[15,16] The basis for this acousto-optic memory correlator is the acousto-photorefractive effect, by which an index-of-refraction pattern corresponding to an rf signal is stored in $LiNbO_3$. Operation is similar to acousto-optic correlators using one fixed reference mask except that the fixed pattern is on the same substrate as the live acoustic signal. Storage is in YZ $LiNbO_3$ and is effected by exposure of the $LiNbO_3$ to a high-intensity, short-duration laser pulse simultaneously with the propagation of a SAW signal in the Z-direction. The laser light may be either 533 nm incident in the X-direction or 1060 nm incident in the Y-direction. The former storage mode succeeded in producing an 8-week storage time for a 7-bit Barker code; storage times for the latter mode are typically a few hours. Signals with center frequencies up to 90 MHz have been successfully stored using a 3-ns storage pulse. The dynamic range of this device is about 25 dB. The T-B product is limited because only a 1-MHz bandwidth has been employed so far. However, there do not appear to be any obstacles to storing at large fractional bandwidths. It has been verified that the storage at 1060 nm is due to electrons trapped on the surface of the $LiNbO_3$. The source of the electrons is most probably photoemission from the $LiNbO_3$. Further work is required to verify the storage mechanism and to elucidate the cause for a fatiguing effect whereby restorage is at a lower efficiency unless sufficient time has passed (~1 day) or the $LiNbO_3$ has been annealed at about 250° C.

The acousto-optic correlator may be used for rapid spectral analysis when it is employed as a building block for performing a Fourier transformer via the chirp transform algorithm.[17] The Fourier transform of a function x(t) is defined as

$$X(f) = \int_{-\infty}^{\infty} X(f) \exp(-j^{2\pi f}t)\, dt \quad . \qquad (3)$$

By substituting the identity $-2ft \equiv (f - t)^2 - f^2 - t^2$ into eq. (3), and rearranging terms, the following is obtained:

$$X(f) = e^{-j\pi f^2} \int_{-\infty}^{\infty} [x(t) \exp(-j\pi t^2)]\, [\exp(j\pi(f - t)^2)]\, dt \quad . \qquad (4)$$

The first bracketed term of the integral may be identified as a premultiplication of the function x(t) by the complex representation of a linear FM chirp. This product is then correlated with a complex chirp, and the result is postmultiplied by a complex chirp. Without the last step of postmultiplication, eq. (4) is recognized as the principle of the microscan receiver which is widely used for rapid spectrum analysis. This architecture can be implemented with an acousto-optic correlator. Thus, spectral analysis can be performed in a time on the order of the propagation time of the chirp signals through the acoustic delay lines--several tens of microseconds at most. During the propagation of the chirps, the frequency components are output as a function of time from the photodetector. With a large time-bandwidth processor, equivalent amounts of spectral data are obtained more rapidly than with the Bragg cell analyzers described in section II. Since coherent detection is possible with the acousto-optic correlator, there is also the advantage of greater dynamic range than with direct light intensity detection. Using a SAW acousto-optic correlator with a T-B product of 1000, an instantaneous bandwidth of 75 MHz and a linear dynamic range of 67 dB have been achieved.[18]

It should be noted that the chirp transform method is more complex than the Bragg-cell analyzer, especially since the transform is performed with a maximum 50-percent duty factor; therefore, two coupled channels are required for 100-percent coverage. Also, this method is theoretically somewhat less capable than the Bragg-cell analyzer in a dense signal environment with multiple simultaneous signals.

The accuracy of the Fourier transform is limited by the accuracy of the broadband chirps required for the large T-B product correlator, as well as by the optics (i.e., optical wavefront distortions). Generally, the chirps are produced by an impulse-excited reflective-array compressive SAW delay line which typically has a 0.1-dB amplitude and 1° phase error.

Since a true Fourier transform is performed, frequency-domain operations may be performed. However, in contrast to the Bragg-cell analyzer, the operations are performed in time rather than in space. For example, excision of a narrow-band interference signal is done by time-gating the photodetector output and reconstructing the time-domain signal obtained by a second chirp-transform operation.

A discrete Fourier transform (DFT) may also be performed acousto-optically using somewhat similar ideas as in the continuous transform. The DFT (denoted by G_k) of a sample sequence $\{g_n\}$ is defined as

$$G_k = \sum_{n=0}^{N-1} g_n \exp\left(\frac{-j2\pi nk}{N}\right), \qquad (5)$$

where n is the total number of samples in the sequence. By substituting the identity

$$nk = [(k + n)^2 - (k - n)^2]$$

into eq. (5) the following equation results:

$$G_k = \sum_{n=0}^{N-1} g_n \exp \frac{-j\pi(k+n)^2}{2N} \exp \frac{j\pi(k-n)^2}{2N}, \qquad (6)$$

where the exponential terms in the equation can be interpreted as being two chirp waves propagating in opposite directions. This architecture for implementing the DFT is referred to as the triple-product convolver (TPC)[19] and can be realized acousto-optically. The sheet beam of laser light used in the acousto-optic convolver (e.g., Fig. 1) for the continuous transform is replaced by a number of (N) discrete, parallel, very narrow laser beams. These beams are intensity-modulated (e.g., by electro-optic modulators) by quadrature signals (for representation of both real and imaginary data), one for each sample g_n. Each beam is detected after it has interacted with two contratravelling chirp waveforms in an acoustic delay line to form the triple product of g_n and the two chirps. The sum of the detected signals contains the real and imaginary parts of the DFT of the input data. The summation of the beam outputs may be obtained either by using N separate photodetectors and recombining the resultant rf outputs or by combining the separate optical beams onto a single detector. The choice of summation method is dictated by convenience or by which method can be made to result in smaller errors. A simple eight-beam acousto-optic TPC has been constructed using a SAW device for the chirp inputs. A real eight-point transform was performed with all g_n = constant. The detected envelope of the output was in very good agreement with the theoretical answer

$$G_k = \frac{8A_o \sin 8kx}{\sin kx}, \qquad (7)$$

where A_o is the amplitude of each laser beam and x is a constant.

One might note that the TPC parallel processes N data samples in the time required for the chirps to propagate through the acoustic delay line. In principle, one might process many data points in several microseconds. In cases where large amounts of data arrive serially at a relatively slow rate, instead of storing the data temporarily to be processed in parallel at the end of the collection time, one could use a modified TPC architecture developed for performing a very long one-dimensional transform.[19] This modified architecture uses charge-coupled device chirp-Z transform modules to modulate the light beams, with proper phasing between light beams. The CCD modules have long interaction time (10^{-1} s) but small bandwidth, while the acousto-optic device has large bandwidth (10^9 Hz) but relatively small interaction time. The combination of these two devices should result in a processor with both large bandwidth and long interaction time.

The TPC architecture appears to be particularly amenable to an integrated-optic implementation. One such implementation is shown in Fig. 2. N laser diodes are coupled to N channel-type optical waveguides on a substrate of $LiNbO_3$ which acts as the SAW medium for the chirps. The waveguides would be produced by titanium in diffusion into the $LiNbO_3$. The laser diode output intensities may be modulated directly by control of the diode current. Each laser beam may be detected individually as shown, or a geodesic lens can be used to focus all the beams to one detector. An investigation into the feasibility of this implementation is presently being carried out. In particular, the effect of the channel waveguides on SAW propagation is being investigated. Further discussion on integrated-optic, acousto-optic devices is given in a paper by Tsai at this symposium.

It should be noted that non-optical techniques already exist to perform the DFT continuous Fourier transforms. The DFT may be calculated using the digital fast Fourier transform (FFT). An eight-bit 10-MHz FFT using parallel architecture has been announced.[20] A TPC architecture has been demonstrated using tapped SAW delay lines.[21] A SAW filter implementation of the chirp transform algorithm (eq. 4) has been demonstrated.[22] A comparison of the various methods for performing the DFT and the chirp transform, both optical and non-optical, is given in Table I. Performance capabilities and the advantages and disadvantages of each method are given.

Time-Integrating Processors

The use of time-integration in signal processors has attracted much recent attention as a method for handling very large time-bandwidth signals such as spread-spectrum signals, which may be of extremely long duration (e.g., days). This method overcomes the limit on T-B product in space-integrating processors due to finite aperture, since it is possible, using present-day photodetector arrays, to integrate for times much longer than delay times in any acoustic cell. It is, however, natural to expect acousto-optic techniques to lend themselves as easily and advantageously to time-integrating processors as to space-integrating processors, since for a moving acoustic signal, time and space coordinates are equivalent variables, related simply by the acoustic velocity. Acousto-optic time-integrating architectures have been demonstrated for correlation, spectral analysis (one- and two-dimensional), ambiguity function processing, and two-dimensional spatial light modulation. Also, various hybrid space- and time-integrating architectures have been demonstrated or proposed.

There are several methods of performing time-integrating correlation. Instead of a space-integrated output of a form such as

$$I(t) \propto \int_{z=0}^{vt} B\left(\frac{z}{v}\right) C\left(2t - \frac{z}{v}\right) \frac{dz}{v}, \qquad (8)$$

appropriate to the convolver of Fig. 1, one now wishes to obtain an output of the form

$$I(z) \propto \int_{t=0}^{T} B(t)C\left(t - \frac{z}{v}\right) dt , \qquad (9)$$

where v is the acoustic velocity. The simplest implementation, by Sprague and Kcliopolous,[23] is to modulate the intensity of a light source with signal B(t). The modulated light illuminates the entire aperture of an acousto-optic cell into which the signal C(t), on a bias, has been introduced. The diffracted light is imaged onto an integrating detector array. The detector output V(t) contains the desired correlation, eq. (9), as well as a DC bias term. Since the extent of Z is limited by the acoustic cell length, one obtains only a finite portion of the correlation function. There is no requirement that coherent light be used. This architecture cannot directly handle frequency-modulated signals.

A second implementation, by Montgomery, uses an unmodulated light source and an acoustic cell having counterpropagating signals B(t) cos $\omega_c t$ and C(t) (cos)$\omega_c t$.[24] The resulting integrated intensity contains DC bias terms plus the correlation integral on a spatial carrier cos $(\omega_c Z/v)$, where ω_c is the temporal carrier frequency of the acoustic signals. By having a spatial carrier, both amplitude and phase may be preserved. These "incoherent" processors illustrate several of the advantages of time-integration. To obtain correlation, there is no requirement to time-invert one of the signals, as in the space-integrating processors. With an unmodulated source, the bandwidth is limited by that of the acoustic cell. The ultimate T-B product achievable is set by the dynamic range of the output detector array.[11]

Time-integrating correlation can also be obtained with "coherent" implementations, where the phase of the light is important. One method uses coherent light with detection at the output array using a coherent reference beam as in an optical interferometer. Light modulation by the acoustic cell is linear with respect to the amplitude rather than the intensity, and the effective bandwidth of the correlation is increased by a factor of two.[25] The correlator by Sprague, et al, in a coherent implementation produces a complex correlation; this correlation is on the amplitude and phase of the spatial carrier that results from a reference beam at an angle to the signal beam. It is also possible to use incoherent light to implement a coherent processor by adding a reference oscillator signal to the acoustic-cell input signals;[20] this implementation has the advantage of a potentially simpler optical system.

A coherent time-integrating correlator using a SAW delay line has been demonstrated.[27] This correlator features optical interference between the Bragg-diffracted beams from each of two counterpropagating SAW signals. One of the attractions of this implementation is that the spatial carrier frequency is the difference between the signal frequency and the center frequency of the SAW device. This spatial frequency is generally much lower than the acoustic frequency and therefore the spatial light pattern may be resolved with a lower-density photodetector array. Both the center frequency and the bandwidth of the signal may be obtained from the spatial light pattern. T-B products exceeding 60 dB have been achieved, and a linear dynamic range of almost 60 dB and a time aperture of 7 μs have been demonstrated. This device has been used to demonstrate the applicability of time-integrating correlators to the processing of spread-spectrum signals. Both direct-sequence and frequency-hopped signals may be detected, and the time-difference-of-arrival of a signal between two separated receiving stations may be determined by the location of correlation peak on the detector array. Bit length in a direct-sequence code and hopping characteristics in a frequency hopper may be obtained with appropriate postprocessing.

A second area of application for the time-integrating processor is in ultra-high-resolution spectral analysis. A chirp transform may be performed using a coherent time-integrating correlator. An extremely long, large bandwidth chirp would theoretically result in high resolution. However, such a chirp is neither practical nor desirable. There is a limit to the bandwidth of the acoustic device, and since the detector array has limited size and element resolution, only a very limited frequency span would be displayed. To minimize the bandwidth requirement, it is desirable to have a small chirp bandwidth and to have the chirp repeated over the duration of the detector integration period or the duration of the input rf signal. A wide rf range will be covered, but this one-dimensional correlator will have the undesirable feature that the output will not be on a continuous display but rather on a comb function, with comb-element width determined by the chirp bandwidth and the number of comb elements equal to the number of times the chirp has been repeated.[25] For this reason time-integrating spectral analysis is done in two dimensional. A second acoustic cell is oriented perpendicular to the first and is used to display all frequencies between adjacent comb frequencies. The output is displayed in two dimensional in a raster pattern approximately orthogonal to the direction of the first acoustic cell.[28] Ultimate frequency resolution is determined by detector integration time. For example, a 30-ms integration time implies a resolution of about 30 Hz.

The two-dimensional spectrum analyzer is actually a specific realization of a much more general processor—the four-product processor. Each acoustic cell may be associated with two input signals. Hence, the same architecture as for the two-dimensional spectrum analyzer can be used to calculate the cross ambiguity function which can be defined as

$$A(\tau,f) = \int_{0}^{T} x(t)y^*(t - \tau)e^{-j2\pi ft} dt, \qquad (10)$$

where the variables τ and f may be interpreted as delay and Doppler, respectively.[26] Calculation of the ambiguity function constitutes a three-product process. Another potential application of the four-product processor is to compensate for Doppler effects which limit ultimate correlation gain in an LPI radar system.[29] The second acoustic cell contains counterpropagating chirps which cancel Doppler shifts on radar return sig-

nals. The correlation peak location also determines the amount of Doppler shift.

Some mention should be made of hybrid space- and time-integrating processors. The simplest is the acousto-optic space-integrating spectrum analyzer whose output falls on an integrating photodetector array.[25] By using low-noise detectors, it is possible to obtain an increase in sensitivity with time-integration by a factor $\sqrt{BT}$, where B is the signal bandwidth at a given detector and T the detector integration time. Other hybrid systems that have been proposed include the use of a space-integrating spectrum analyzer in conjunction with a fixed photographic reference mask for correlation of frequency-hopped spread-spectrum signals.[30]

IV. Summary

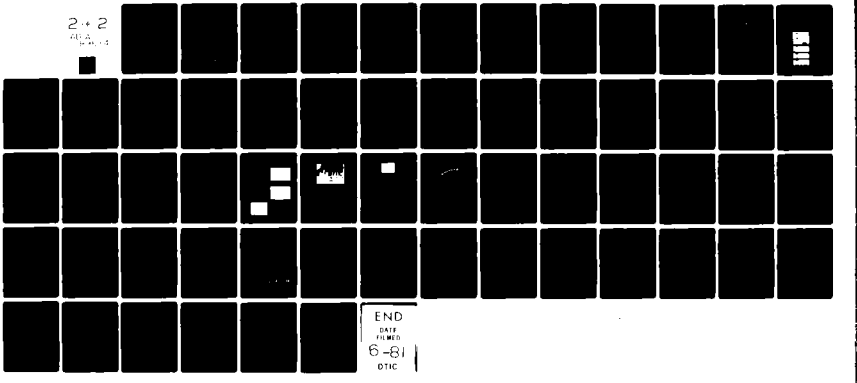
Acousto-optic techniques have been applied to an ever-increasing variety of signal-processing tasks. This has occurred in spite of the fact that we may be reaching ultimate limits in the performance of acousto-optic deflectors. Improved performance in acousto-optic processors has been obtained in a number of ways. Architectures such as the chirp-transform algorithm in acousto-optic implementations can overcome limits on detector read-out time that occur with Bragg-cell spectrum analyzers, while maintaining advantages such as large bandwidth and large linear dynamic range. Acousto-optic time-integrating processors provide time-bandwidth products, or spectral resolution, well in excess of the time-bandwidth product of the acoustic device. In addition to spectral analysis, acousto-optic processors are being applied to tasks such as ambiguity function processing, spread-spectrum signal processing, and two-dimensional spatial light modulation.

References:

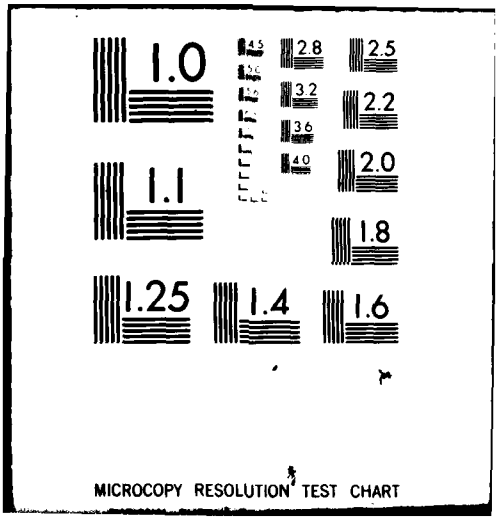
1. A. Korpel, "Acousto-Optics", Appl. Solid State Science, 3, 71-180 (1972).
2. D. L. Hecht, Proc. IEEE Ultrasonics Symp. (1977), 77 CH 1264-1SU, p. 721.
3. J. N. Lee, N. J. Berg, B. J. Udelson, Proc. IEEE Ultrasonics Symp. (1977), 77 CH 1264-1SU, p. 451.
4. N. Uchida and N. Niizek, Proc. IEEE, 61, 1073 (1973).
5. A. Korpel, R. Adler, P. Desmares, and W. Watson, Proc. IEEE, 54 (1966), 1429.
6. C. S. Tsai, L. T. Nguyen, S. K. Yao, and M. A. Alhaider, Appl. Phys. Lett., 26 (1975), 140.
7. J. P. Lee, "Signal Analysis Using the Acousto-Optic Spectrum Analyzer", Defense Research Establishment Ottawa, Technical Note (Oct. 1979).
8. I. Deyhimy, R. C. Eden, R. J. Anderson, and J. S. Harris, Jr., Appl. Phys. Lett., 36 (1980), 151.
9. Litton Amecon, Inc., private communication.
10. D. L. Hecht, IEEE Trans. on Sonics and Ultrasonics, SU-24 (1977), 7. intermod
11. E. B. Felstead, IEEE Trans. on Aerospace and Electronic Systems, AES-3, 907-914 (1967).
12. N. J. Berg, B. J. Udelson, J. N. Lee, and E. Katzen, Appl. Phys. Lett, 32 (1978), 85.
13. R. A. Sprague, Opt. Eng., 16, 467 (1977).
14. M. I. Skolnick, Introduction to Radar Systems, McGraw-Hill, New York (1962), Ch. 9.
15. N. J. Berg, B. J. Udelson, and J. N. Lee, Appl. Phys. Lett, 31 (1977), 555.
16. J. N. Lee, N. J. Berg, and P. S. Brody, Proc. IEEE Ultrasonics Symp. (1979), 79 CH 1482-9, p. 81.
17. L. Mertz, Transformations in Optics, John Wiley & Sons, New York (1965).
18. N. J. Berg, J. N. Lee, M. W. Casseday, and E. Katzen, Appl. Phys. Lett., 34 (1979), 125.
19. J. M. Speiser and H. J. Whitehouse, NUCTN 1355R, Naval Ocean Systems Center.
20. Probe Systems, Inc., private communication.
21. T. M. Reeder and J. L. Swindal, Tech. Report ECOM-73-0194-F (Nov. 1974).
22. H. M. Gerard, P. S. Yao, and O. W. Otto, Proc. IEEE Ultrasonics Symp. (1977), 77 CH 1264-1SU, p. 947
23. R. A. Sprague and C. L. Koliopolous, Appl. Opt., 15 (1976), 89.
24. R. M. Montgomery, U. S. Patent No. 3,634,749 (1972).
25. T. M. Turpin, Proc. of SPIE, 154 (1978), 196.
26. P. Kellman, Proc. of SPIE, 214 (1979), 63.
27. N. J. Berg, I. J. Abramovitz, J. N. Lee, and M. W. Casseday, Appl. Phys. Lett., 36 (1980), 256.
28. C. E. Thomas, Appl. Optics, 5 (1966), 1782.
29. N. J. Berg, I. J. Abramovitz, J. N. Lee and M. W. Casseday, Acousto-Optic Time Integrating Correlator for Detection and Characterization of Broad Band LPI Communications, to be published in Proc. 1980 Army Science Conference.
30. D. Psaltis and D. Casasent, Appl. Optics, 18 (1979), 3203.

AD-A099 614 TEXAS TECH UNIV LUBBOCK DEPT OF ELECTRICAL ENGINEERING F/G 9/4
WORKSHOP ON FUTURE DIRECTIONS FOR OPTICAL INFORMATION PROCESSIN--ETC(U)
MAR 81 J F WALKUP, T F KRILE DAAG29-80-C-0110

UNCLASSIFIED ARO-17468.1-EL NL



END
DATE FILMED
6-81
DTIC


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6
MICROCOPY RESOLUTION TEST CHART

Function	Technologies	Characteristics	Applications	Advantage	Disadvantage
Bragg cell	Bulk wave and surface wave acousto-optic (a/o) integrated-optic spectrum analyzer	1-GHz bandwidth. 1-MHz resolution. 50-dB dynamic range	Spectrum analysis. cw excision	100-percent coverage	Relatively slow readout
Chirp transform microscan	Surface acoustic wave (SAW) reflective array compressor. a/o convolver	~100-MHz bandwidth. 50-kHz resolution. 5-bit accuracy. 65-dB dynamic range	Spectrum analysis. cw excision	High-resolution. rapid processing	Complex
	Time integrating a/o correlator	1-MHz bandwidth. 1-Hz resolution. 3-bit accuracy. 45-dB dynamic range	Spectrum analysis	Ultra-high resolution	Limited dynamic range
Discrete Fourier transform	Digital fast Fourier frequency	256-kilo point. 10-MHz bandwidth. 8-bit accuracy	Spectrum analysis. cw excision. complex arithmetic	High accuracy	Limited bandwidth
	Hybrid SAW	10-Megabit/s. 8-bit accuracy. 45-dB dynamic range		Large bandwidth	Complex
	Triple-product convolver (a/o convolver)	1-MHz bandwidth. 8-bit accuracy. 60-dB dynamic range	Spectrum analysis cw excision. complex arithmetic and beam forming	Large bandwidth high data throughput	Complex

Table I. Fourier Transform Techniques

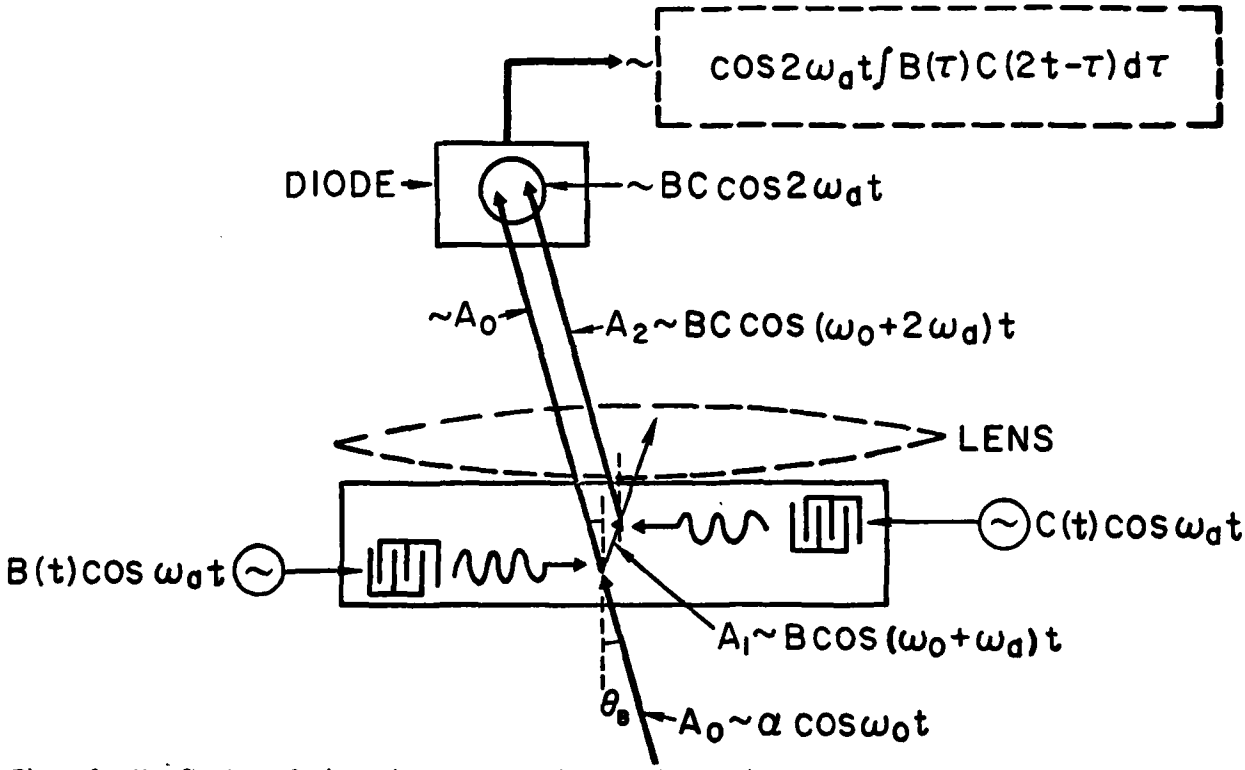


Figure 1. Waveform convolution using accousto-optic Bragg interaction.

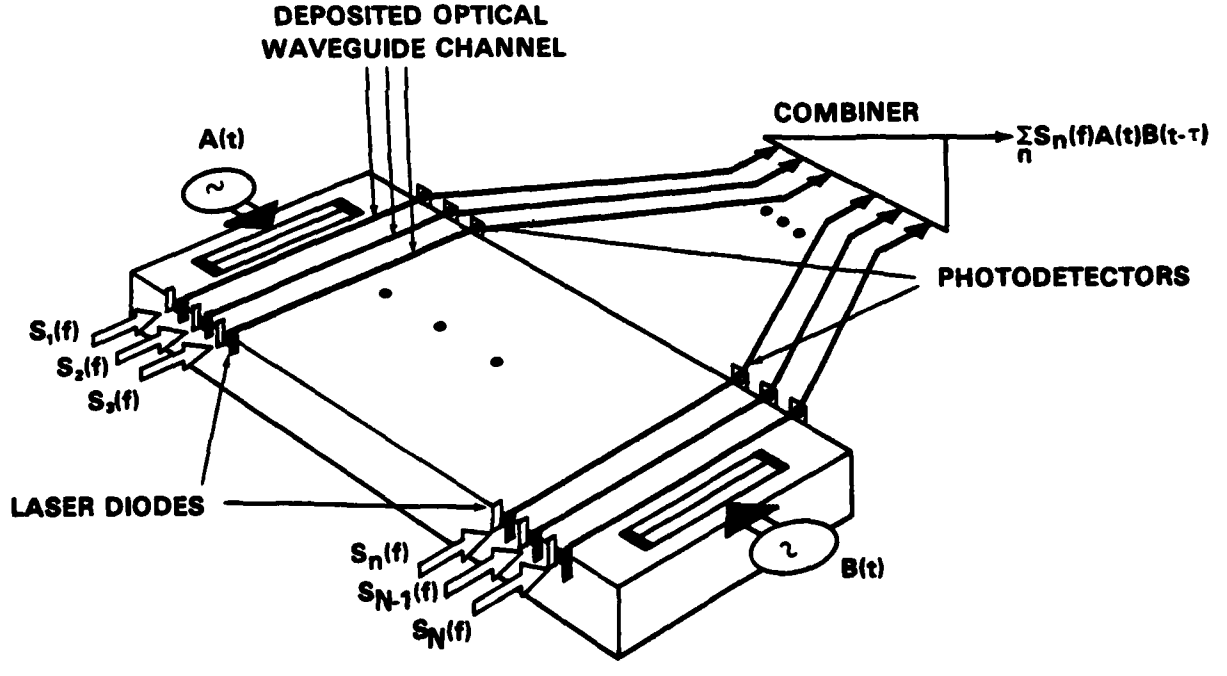


Figure 2. Potential integrated-optic implementation of the triple-product convolver.

Discussion (John N. Lee; Discussion Leader: Harper Whitehouse)

Q. In your discussion about the detector plane, you talked about GaAs. Why GaAs instead of silicon, which we already have the technology for?

A. High mobility. The speed at which you can shift out information from a shift register depends on how fast you can move charge packets from one element to the next. GaAs, with its high mobility, is the best material right now.

Q. In your discussion of the acousto-photorefractive effect, you postulated having some external photo emission in order to record charge packets in the array correlator. That's a little disturbing because L_iNbO_3 has a pretty wide bandgap, about 3.5 or 4 ev., and a work function of 3.5 or 4 ev., so that's a four photon process to actually emit something. I'm wondering if that's a standing acoustic wave.

A. Yes, it is.

R. But even then the YAG laser is taking a 20 nanosecond shot, so I wonder if that's not just the transverse surface acoustic field driving the standard kind of diffusion drift photo-refractive effect temporarily during the 20 nanosecond onset of the laser, and not an external photoemission. The surface charge of the L_iNbO_3 substrate should change astronomically during the photoemission process, so all you have to do to test the hypothesis is to measure the surface voltage and see if it actually changes.

A. The reason we came to the conclusion that it was possibly photoemission was that, even without the SAW present, when you illuminate the device with the laser, it turns positive. The surface has a positive charge pattern on it corresponding to the laser beam, and this occurs on both sides of the crystal. It doesn't look like it is simple charge transport because we can't find a correspondingly negatively charged region for every positively charged region, in fact there are no negatively charged regions.

Q. Could you comment on the relative merits of using bulk crystal devices vs. thin film devices?

A. Bulk wave devices have been around a lot longer and generally they can operate at higher frequencies because of the way transducers are made. They are thin film transducers, and you can go to higher frequencies by making them thinner. The limitation on the SAW is how fine you can make the finger pattern. You have interdigital transducers in the SAW case, and eventually you can't make the lines fine enough and good enough so that you can get a good transducer. That limits you generally to about 1 GHz center frequency whereas the L_iNbO_3 bulk wave lines can go up to about 2.5 GHz. They're working directly in L-band right now. However, with the planar technology of SAWs you can do a lot of nice things, like suppress triple transit, get flat bandwidths, do phase-matching, get uniphase wavefronts. These are some things that are different about the two.

Q. Can you elaborate on that? It seems that lithography has been making great strides. People claim that with E-Beam technology and X-ray lithography, they can make very fine lines. Do you see this problem with defining transducers as an ultimate limitation? Or is it a temporary problem that will be resolved with more work?

A. To a certain extent, it is temporary. We have not done any E-beam lithography, but it should be possible. One problem you have to watch for is that of matching the impedance of the transducer to your source. When you start changing dimensions, you have to watch how the impedance changes, and you have to be able to match out the interfinger capacitances properly in order to get the lowest insertion loss. I've seen approaches where people used bulk wave transducers for their RF input and they bounce the bulk wave up against a grating to convert it into surface waves simply because they can make the grating much better than they can make a true transducer.

R. I think you're right that progress made with X-ray and E-beam technology would enable us to fabricate SAW transducers with sub-micron dimensions, which would mean center frequencies of 4 to 5 GHz. But at this time there is no need to push it further, because attenuation losses are too high for higher frequencies. If we find low acoustic loss materials for SAW or bulk waves, then there will be a need for fabricating transducers at higher frequencies, and we are confident that by that time progress made in E-beam and X-ray lithography will enable us to do the fabrication.

Q. You gave a good summary of the state of acousto-optics. What do you think are the important areas that we should be addressing in order to bring these devices to fruition?

A. I would say the post-processing. The detectors need improvement, and you need to find a way to improve the processing speed, that is what holds up the whole works.

Q. In a coherent processor, for example a spectrum analyzer where you are trying to determine magnitude and phase, your detector elements must be packed together a good deal closer than in an incoherent processor. How much of a penalty do you pay there?

A. As you go farther off the center frequency on our device the fringe pattern gets finer. Eventually you are putting two fringes onto one detector element, and then you lose any hope of getting information out.

With detector elements 13 microns across, 25 micron center-to-center spacing, and 1 to 1 imaging of the acoustic cell onto the detector array, you get a 64 MHz frequency span. However, if you are willing to tolerate a smaller time window, you can expand the beam, or conversely you can make a longer array.

Q. If you are looking for just magnitude, what sort of bandwidth would the processor have, all other things being the same?

A. I would say it is the bandwidth of your acoustic cell, so you could be paying quite a penalty in order to get phase information.

Optical Processing Using Integrated Optics*

Chen S. Tsai
Department of Electrical Engineering
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

I. Introduction-Guided Wave and Integrated Optics Technology

Integrated Optics is an emerging technology that has been actively pursued in recent year. The ultimate objective of integrated optics is to realize miniature optical components such as light sources, modulators, switches, deflectors, lenses, prisms, and detectors in a substrate to perform various useful functions. Like the existing integrated electronics in which a large number of active and passive components such as transistors, diodes, resistors and capacitors are packed in a small area semiconductor chip, the integrated optics, when fully developed, are expected to have many advantages over the existing bulk optical systems. Some of the advantages of the miniature components over their bulk counterparts are smaller size and lighter weight, wider bandwidth, lesser electrical drive power requirement, greater signal accessibility and integratability. The miniature components are also expected to possess advantages in stability, reliability, ruggedness and ultimate cost.

It has now been well recognized that the most immediate and important applications of integrated optics lie in the areas of wideband multichannel communications (for both military and civilian) and signal processings (for military hardwares such as radars). With regard to communications, a number of firstgeneration low data rate laboratory and field test systems have been built and their performance have demonstrated their potential. With regard to signal processings, research and development activities have increased considerably as a result of recent progress in guided-wave acoustooptics. (1,2) For example, a number of military and civilian laboratories are currently using wideband planar acoustooptic deflectors to develop an integrated optic spectrum analyzer for data processing of very wideband radio frequency signals (3.4). In addition. recent progress on guided-wave electrooptic devices including analog to digital converters,(5-7) directional coupler switches (8), total internal reflection (TIR) switches (9, 10) and networks (10-12), and bistable switches (13) is a welcome news to optical information processing community.

In this paper a brief review of the most recent progress on planar waveguide acoustooptics, channel waveguide acoustooptics, and channel waveguide TIR switching network is made. Some related topics for future research are also included.

II. Planar Waveguide Acoustooptics

A. Wideband Bragg Modulators and Deflectors

The basic configuration of planar waveguide acoustooptic (AO) Bragg modulators and deflectors is shown in Fig. 1. The deflection angle and the intensity of the Bragg diffracted light are, respectively, proportional to the frequency and the power of the RF or the acoustic signal applied to the surface acoustic wave (SAW) transducer. It is now possible to design and fabricate high-performance planar acoustooptic (AO) Bragg modulators and deflectors with a bandwidth approaching one GHz.(2). For example, the deflectors of 700 and 470 MHz bandwidth have been realized most recently in Y-cut $LiNbO_3$ waveguides using multiple tilted transducers and a tilted-finger chirp transducer,[14,15] respectively With some improvement in transduction efficiency of the SAW transducers an ultimate diffraction efficiency of about 50% at one watt RF drive power for one GHz bandwidth can be expected.[2] The measured efficiency of the 700 MHz and 470 MHz bandwidth deflectors referred to above are 10% at one watt RF drive power and 16% at 0.2 watt RF drive power, respectively. Finally, an ultimate time-bandwidth product (defined as the product of the acoustic transit time across the incident light beam aperture and the modulator/deflector bandwidth) of up of 4000 may be achievable (2).

B. Integrated Acoustooptical Circuits. Integrated AO circuits require integration of all active and passive optical components on a *single* substrate or a small number of substrates. Using an RF spectrum analyzer as an example, a fully-integrated or monolithic AO circuit is depicted in Fig. 2. The AlGaAs multilayer structure is a potential substrate for this monolithic module. In this case the light source takes the convenient form of a distributed feedback or Bragg reflector laser. However, most of the other passive and active components for this monolithic module remain to be developed. At present, silicon (thermally oxidized) and lithium niobate are the two most promising substrates for implementation of a hybrid module.[3] In the former the light source such as a GaAlAs DH laser diode is butt-coupled (17) to one edge of the Si substrate. In the latter both the GaAlAs DH laser diode and the photodetector array are butt-coupled to the edges of the $LiNbO_3$ substrate (see Fig. 3). The wideband AO modulator/deflector together with recent progress on fabrication of miniature laser sources, waveguide geodesic lenses, and photodetector arrays, integration of all or most of these components on a common substrate is becoming a reality.

A model design has shown that a Y-cut $LiNbO_3$ plate having a substrate area of 2.5x8 cm is sufficient to accommodate all passive and active components. In view of the fact that high-quality optical waveguides, geodesic lenses, and wideband high-efficiency Bragg deflectors/modulators have been successfully fabricated in the $LiNbO_3$ substrate, the hybrid structure as shown in Fig.3 appears to be the most attractive approach

*This work was supported by the AFOSR, AROD, NSF, and BMDATC.

for the present. Implementation of this hybrid AO circuits is being carried out at a number of research laboratories in the United States.

C. Some Potential Applications. Clearly, the resulting integrated AO modules or circuits should find a number of unique applications in wideband multichannel optical communications and RF signal processing. As indicated earlier, one application that has already received a great deal of attention and interest is real-time spectral analysis of wideband rf signals. Other applications, such as correlation, convolution, pulse compression and matched filtering of rf signals, and multiport switching of optical signals in a single-mode fiber optic communication system, (2) should not be far behind.

One of the most important classes of optical signal processors is based on the use of coherent acoustooptic interactions. A number of bulk-type AO signal processors have been studied and demonstrated. (18) Most of these bulk-type and quasi bulk-type AO processors for one-dimensional signal processing may be implemented using the planar guided-wave structure. In fact, some encouraging results have already been demonstrated in the experiments on spectral analysis, (1,2) convolution, (1,2) pulse compression, (19) and time-integrating correlations (20). Better results can be expected using the integrated AO circuits. (21) In this subsection two of the most promising applications, namely, spectral analysis and time-integrating correlation are decribed.

1. Spectral Analysis of Very Wideband RF Signals. As mentioned in the Introduction, implementation of an RF spectrum analyzer using monolithic or hybrid integrated optic technique has already received a great deal of attention and interest in both military and civilian communities. Fig. 2 depicts a schematic diagram of a monolithic integrated optic spectrum analyzer. When a spectrum of rf signals are applied to the transducer, each spectral component generates a SAW which deflects the incident light beam in a corresponding direction. As indicated previously, the deflection angle and the intensity of the Bragg diffracted light are, respectively, proportional to the frequency and the power of the RF or the acoustic signal. Thus, by measuring the linear positions and the intensities of the deflected (diffracted) light spots at the focal plane of the transform lens the power spectral density of the rf signal of interest may be determined. In an actual spectrum analyzer, the Bragg diffracted light beams are focused on a photodetector array and the rf spectra is read out using a CCD array. The number of resolvable channels and the corresponding frequency resolution are given in Ref. 2. For the 680 MHz bandwidth deflector which we have fabricated most recently in a Y-cut $LiNbO_3$ waveguide, (15) the measured frequency resolution is 0.6 MHz for a truncated Gaussian light beam of 6 mm aperture. Thus, based on the measured resolution this deflector would provide 1130 resolvable frequency channels. This deflector has also provided the deflected and undeflected light spots of very fine quality.

The integrated optic RF spectrum analyzers, when fully developed, are expected to possess two major advantages: 1. increased performance and reduced cost over both currently employed technology and competing technologies, and 2. reduced size and increased compactness.

2. Time-Integrating Correlation of Wideband RF Signals. Acoustooptic time-integrating correlators (AOTIC) perform correlation by using a closely spaced photodetector array to integrate in time for each point within the Bragg cell. (22,23) A fully-integrated or monolithic guided-wave version (2) is depicted in Fig. 4. The signal to be correlated, $S_1(t)$, is added with a bias voltabe V_1 and used to modulate the intensity of a wave-guide laser source. The modulated light is then collimated and diffracted by the surface acoustic wave produced by an RF carrier which is amplitude-modulated by the reference signal $S_2(t)$. A proper choice of bias voltage V_2 would ensure that the intensity of the diffracted light is linearly proportional to $S_2(t)$ The diffracted light is then collected by a waveguide lens, filtered, and imaged onto a photodetector array. It can be shown that, if $S_1(t)$ and $S_2(t)$ have zero means values, the intensity of the diffracted light at the output of the photodetector array contains the correlation signal between $S_1(t)$ and $S_2(t)$. This correlation signal is displayed in space but can be read out in time using a CCD array. Since the correlation is performed in time rather than in space this type of correlator is potentially capable of a very long processing time which is determined by the time constant of the photodetector array. Furthermore, since both the coherent light source and the AO Bragg cell can be modulated at GHz bandwidth, this type of correlator is also potentially capable of very large bandwidth, and thus very large time-bandwidth product. Our preliminary experiments using surface acoustic waves (centered at 125 MHz) in a Y-cut $LiNbO_3$ waveguide and a He-Ne laser light (6328 A) have demonstrated a processing time of 7 milliseconds and a time-bandwidth product of 1.5×10^5 (20).

As in the spectrum analyzers, although the AlGaAs multilayer structure is the ideal substrate for the monotithic AOTIC modules the hybrid structure using a $LiNbO_3$ substrate, as illustrated in Fig. 5 constitutes the most attractive and realizable approach for the present. A model design has shown that a Y-cut $LiNbO_3$ plate having a substrate area of 2.5 x 6 cm is sufficient to accommodate all passive and active components. Based on the aforementioned preliminary results, (21)this hybrid AOTIC should be capable of providing much better performance figures (21).

III. Channel Waveguide Acoustooptics

As indicated in Section II, planar AO interactions and devices have already been shown highly useful for wideband multichannel integrated optic communication and signal processing systems. (2) On the other hand acoustooptic Bragg deflection in channel waveguides in which the optical waves are confined in the channels has received much less attention heretofore. Nevertheless, the resulting channel devices are potentially more useful in fiber optic systems because of the compatibility in dimension and, thus, the relative simplicity in facilitating the coupling between the channel waveguide and the optical fiber.

One interaction configuration of great interest is shown in Fig. 6. Two identical channel waveguides are crossed at an angle Ψ. One unique characteristic of these crossed guides is that the refractive index

change in the cross-over (intersection) region is twice that in the channel. As a result, the crosstalk between the two guides can be made small. (10,24) An interdigital transducer is symmetrically positioned so that the SAW generated propagates in the intersection region. The center frequency of the SAW is such that the corresponding Bragg angle is equal to one half of the intersection angle. An optical wave incident at guide 1 is Bragg diffracted by the moving optical grating induced by the SAW. Consequently, a portion of the incident light is deflected to guide 2. The frequency of the deflected light is up-shifted by an amount equals to the acoustic frequency. Similarly, an optical wave incident at guide 2 will have portion of its intensity deflected to guide 1 and have the frequency of the deflected light down-shifted by the same amount. Such a device configuration will have a variety of unique applications in future integrated and fiber optic systems such as double-pole-double-throw switching, time-division multiplexing, and demultiplexing, and heterodyne detection. In the last application the frequency-shifted light can be conveniently used as a reference signal (local oscillator) in connection with optical communications and fiber optic sensing.

We have recently demonstrated encouraging results with the device configuration just described in a Y-cut $LiNbO_3$ substrate. (25) The center frequency of the SAW is 634 MHz, appropriate for the 3.0^o intersection angle between guide 1 and guide 4. Ten finger pairs of 0.77 mm aperture are used in the transducer. A diffraction efficiency of up to 67% has been obtained using an rf power of 800 mw. Since the measured conversion efficiency of the transducer is -12 db and a conversion efficiency of -6 db had been routinely demonstrated with this type of transducer in this laboratory, it should be possible to reduce this drive power by a factor of 4. A -3 db deflector bandwidth of 71 MHz has been measured. Finally, when acting as an optical switch a switching time (defined as the time between 0 and 100% points) of 25 nanoseconds has been measured. (25)

IV. Optical Switching Networks and Matrices Using TIR Switches

A. Improved Optical Channel Waveguide TIR Switch.

We have recently demonstrated some superior features of an improved version of the TIR switch (10) which does not require any horn section (Fig. 7). This switch employs straight $2\Delta n$ intersections in a Y-cut $LiNbO_3$ substrate. As indicated previously in Section III, the crosstalk of this type of intersections can be made small.

The $2\Delta n$ intersections of various intersecting angle θ were first formed in a Y-cut $LiNbO_3$ substrate by the now well-established Ti-diffusion method. The width of each channel waveguide is typically 20 um. A pair of parallel metal electrodes having 5 um in separation and typically 1.6 mm in length was then deposited at the center of the intersection region. Switching and modulation experiments were carried out using a 6328 Å He-Ne laser light propagating in the TE fundamental mode. Performance figures considerably better than that obtained with the device with horn sections (9) have been measured.(10) For example, the drive voltages of only 5, 8, and 11 volts are, respectively, required for maximum switching in the switches with 1.0^o, 1.5^o, and 2.0^o intersections. Fig. 8 shows the modulation curves for the switch with the intersection angle of 1.0^o. The corresponding switching efficiency in light power are 93, 88, and 92%. The corresponding crosstalks or extinction ratio which is defined as the ratio of the switched (reflected) light power and the unswitched (transmitted) light power in db are measured to be 15.7, 17.5, and 17.0 db. The measured insertion loss is around 1.3 db for all switches. This loss is mainly caused by the electrodes and should become smaller if a dielectric buffer layer is added. Thus, it is reasonable to suggest that higher switching efficiencies and smaller crosstalks than those just mentioned should be achievable after such a buffer layer has been added to the switches. Finally, the calculated capacitances of the electrode pairs indicate that the corresponding base bandwidths of the three switches with a 50 ohm termination should be 5.9, 7.8, and 11.8 GHz, respectively. In view of the low drive voltage requirement and the subnanosecond switching speed, and the simplicity in device design/fabrication this improved TIR switch should find a variety of applications including the residue arithmetics based optical computers (26).

B. 4x4 Optical Switching Network/Matrix. High-speed low-drive power optical switching networks and matrices should perform important functions in future integrated and fiber optic systems. Examples range from pure optical systems such as multichannel fiber optic communication terminals (27) and optical computers (26) to hybrid systems such as electronic computer networks and communication buses. An integrated optical switching matrix (network) was first described a few years ago, (11) and a 4x4 switching matrix was later realized in a $LiNbO_3$ substrate (12) using five stepped $\Delta\beta$ reversal directional-coupler switches. (8) In view of the simplicity and the superior performance of the improved TIR switch just described in the last subsection we have most recently fabricated a 3x3 and a 4x4 switching networks consisting of such switched in Y-cut $LiNbO_3$ substrates. Note that 3x3 and 4x4 switching networks/matrices are the basic building blocks for large networks/matrices.

A 4x4 switching network which consists of 4 straight channel waveguides and 5 improved TIR switches is shown in Fig. 9. It is clear that an optical signal entering at any of the input ports can be routed to any of the output ports by proper combination of the individual switches being activated. The dimensions of the 4x4 switch we have experimented with are 6.18 mm x 0.20 mm in which 6.18 mm is the total length of the switch. The width of each channel waveguide is 20 um. Very encouraging results have been obtained with this preliminary matrix switch. For example, Fig. 10 shows the waveforms detected at the four output ports when the light enters the second input port (I_2) and both Switch S_2 and Switch S_5 are activated. Note that the input light is a cw He-Ne laser at 0.6328 um. We see that the input light is routed to output port O_2 with very high efficiency ($\sim$ 95%) and the crosstalks detected at output ports O_3 and O_4 are low ($\sim$ -17 db). The preliminary results just mentioned are very significant in view of the simplicity of the switching network and the various advantages that can be expected from the basic TIR switch.

V. Some Topics Suggested for Future Research

The recent progress in guided-wave acoustooptic and electrooptic devices as decribed above has suggested a number of related topics for further research. Some of these topics are as follows:

1. Study of alternate substrate and AO interaction media: Growth and characterization of thin-film materials which have higher AO figure of merit than $LiNbO_3$ such as As_2S_3, TeO_2 and Te are worth undertaking. A higher AO figure of merit implies a lower acoustic power density and, thus, a lower acoustic nonlinearity and a higher dynamic range. (2)
2. Search for better methods for efficient electrical to acoustical transduction in conjunction with silicon and GaAs substrates: Methods which are simpler and/or more efficient than the commonly used ZnO overlayer (28) are highly desirable.
3. Implementation of one or several integrated optic modules or circuits to identify basic and important problem areas for in-depth research: The Navy-Air Force program on integrated optic spectrum analyzers may serve as a model for this undertaking. Integrated optic switching network or matrix is one of the potential candidates for this purpose.
4. Identify and study suitable signal processing architectures in conjunction with the integrated AO circuits: The schemes capable of multi-dimensional time-integrating correlations are of particular interest.
5. Study of new opto-acoustic and opto-electronic interactions: Study of high-risk but basic interactions and phenomenon deserves support. This type of basic research may open up new types of device and application. One example of this type of research is the efficient diffraction of submillimeter wave from free-electron density waves in piezoelectric semiconductors such as n-type InSb reported recently.(28)

References

1. a. C.S. Tsai, Le T. Nguyen, B. Kim, and I.W. Yao, "Guided-Wave Acoustooptic Signal Processors for Wideband Radar Systems," SPIE, Vol. 128, Effective Utilization of Optics in Radar Systems, pp. 68-74, February 1978.

   b. C.S. Tsai, "Surface Acoustooptic Devices - Fundamentals and Wideband Applications," Proc. of Workshop on Applications of Integrated Optics to Missile Guidance, pp. 16-1 to -18, Edited by B.D. Guenther and C.M. Verber, April 18-19, 1978, Redstone Arsenal, Alabama.

2. C.S. Tsai, "Guided-Wave Acoustooptic Bragg Modulators for Wideband Integrated Optic Communications and Signal Processings," IEEE Trans. on Circuits and Systems, Special Issue on Integrated and Guided Wave Optical Circuits and Systems, Vol. CAS 26, 1072-1098 (Dec. 1979); Edited by H.F. Taylor.
3. M.C. Hamilton, D.A. Wille, and W.J. Micceli, "An Integrated Optical RF Spectrum Analyzer," Proc. 1976 Ultrasonics Symp., IEEE Cat.
4. T.G. Giallorenzi, presented at 1978 Topical Meeting on Integrated and Guided-Wave Optics, Technical Digest, pp. Ma 1-1 to -3, OSA 78CH1280-7QEA.
5. S. Wright, I.M. Mason, and M.G.F. Wilson, Electron. Lett., Vol. 10, 508 (Nov. 1974).
6. H.F. Taylor, Proc. IEEE, Vol. 63, 1524 (Oct. 1975); D. Lewis and H.F. Taylor, "An Integrated Optical Analog-To-Digital Converter," Presented at AGARD Meeting, May 16-20, 1977, London, England.
7. P. Saunier, C.S. Tsai, and I.W. Yao, presented at 1978 Topical Meeting on Integrated- and Guided-Wave Optics, Salt Lake City, Utah, Technical Digest, pp. TuC2-1 to -4, OSA 78CH1280-7 QEA.
8. H. Kogelnik and R.V. Schmidt, IEEE J. Quantum Electron., QE-12, 396 (1976).
9. C.S. Tsai, B. Kim, and F. El-Akkari, IEEE J. Quantum Electron., Vol. QE-14, 513 (1978).
10. F. El-Akkari, C.L. Chang, and C.S. Tsai, presented at 1980 Topical Meeting on Integrated- and Guided-Wave Optics, Jan. 28-30, Incline Village, Nevada, Technical Digest, pp. TuE4-1 to -4.
11. H.F. Taylor, Electron. Lett., Vol. 10, 41 (1974).
12. R.V. Schmidt, Electron. Lett., Vol. 12, 575 (1976).
13. See, for example, Workshop on Optical Bistable Devices, Sponsored by U.S. Army Research Office, May, 1980.
14. C.S. Tsai, M.A. Alhaider, Le T. Nguyen, and B. Kim, Proc. IEEE, 64, 318 (March 1976).
15. C.C. Lee, K.Y. Liao, C.L. Chang, and C.S. Tsai, IEEE J. Quantum Electron., Vol. QE-15, 1166 (October 1979).
16. K.Y. Liao, C.L. Chang, C.C. Lee, and C.S. Tsai, Proc. of 1979 Ultrasonics Symposium, pp. 24-27, IEEE Cat. No. 79CH1482-9SU.
17. R.G. Hunsperger, A. Yariv, and A. Lee, Appl. Opt., Vol. 16, 1026 (April 1977).
18. a. See, for example, R.W. Damon, W.T. Maloney, and D.H. MaMahon, Physical Acoustics, Ed. by W.D. Mason and R.N. Thurston, Academic Press, New York, Vol. 7, 273 (1970) and Proceedings of Bulk-Wave Acoustooptics Workshop, Nov. 28-30, 1979, Monterey, California, SPIE, Vol. 214 (1980).

   b. J. Lee and N. Berg, Presented at Conference/Workshop on Acoustooptic Bulk-Wave Devices, Nov. 27-29, 1979, Naval Post Graduate School, Monterey, California., SPIE, Vol. 214 (1980).

   c. IEEE Proceedings, Special Issue on Acousto-Optic Signal Processing, Jan. 1981; Editor, A. Vanderlugt.
19. C.S. Tsai, Le T. Nguyen and C.C. Lee, 1978 Topical Meeting on Integrated and Guided-Wave Optics, Jan. 16-18, Salt Lake City, Utah, Technical Digest, pp. TuC3-1 to -4.
20. I.W. Yao and C.S. Tsai, 1978 Ultrasonic Symposium Proceedings, IEEE Cat. No. 78CH1344-1SU, p. 87.
21. C.S. Tsai, J.K. Wang, and K.Y. Liao, SPIE, Vol. 180, 160 (1979).
22. R.A. Sprague and K.L. Koliopaulos, Appl. Opt., Vol. 15, 89 (1976).
23. T.M. Turpin, SPIE, Vol. 154, 196 (1978).
24. T. Kurokawa and S. Oikawa, Appl. Opt., Vol. 16, 1033 (1977).
25. C.S. Tsai et al., presented at 1980 Topical Meeting on Integrated- and Guided-Wave Optics, Jan. 28-30 Incline Village, Nevada, Technical Digest of Post-Deadline Papers, pp. PD7-1 to -4.

26. A. Huang, Y. Tsunoda, and J.W. Goodman, Appl. Opt., Vol. 18, 149 (January 1979).
27. H. Wichansky and L. Dworkin, Proc. of 1978 NSF Meeting on Optical Communication Systems, pp. 192-197, June 5-7, Pittsburgh, Pa.
28. D. Mergerian et al, Proc. 1978 Ultrasonics Sym., IEEE Cat. 78-CH1344-1SU, pp. 64-69.
29. V.V. Proklov et al, Presented at 1979 International Conference on Solid State Devices, Aug. 27-29. Tokyo, Japan.

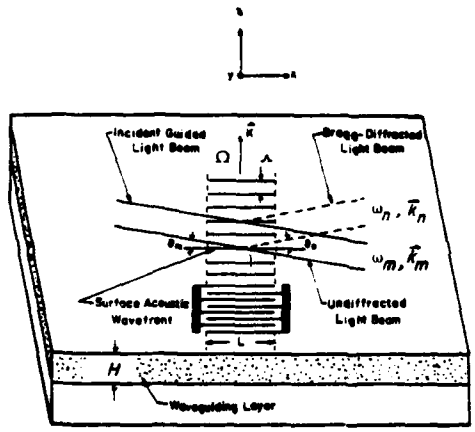


Fig.1 Guided-Wave Acoustooptic Bragg Diffraction From A Single Surface Acoustic Wave

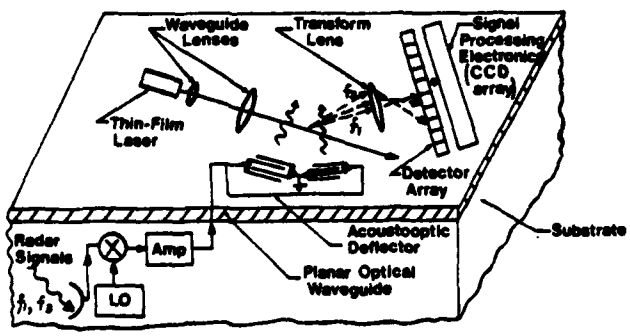


FIG.2 Integrated Optic Spectrum Analyzer

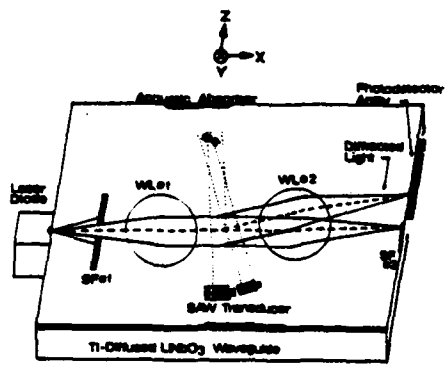


FIG.3 An Acoustooptic Spectrum Analyzer Using Hybrid Optical Waveguide Structure

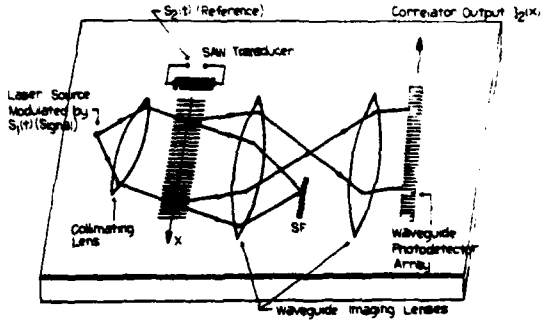


FIG. 4 GUIDED-WAVE TIME-INTEGRATING ACOUSTOOPTIC CORRELATOR

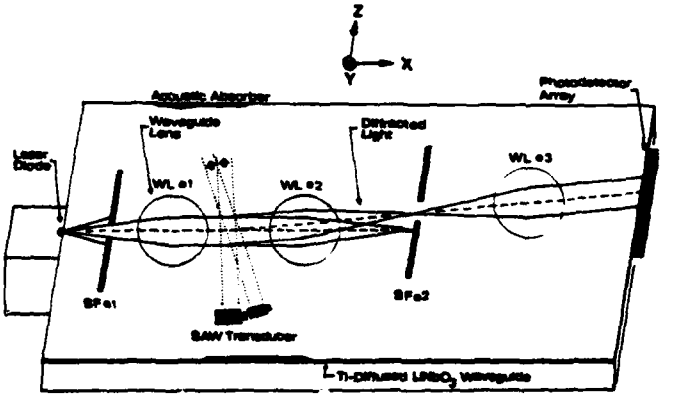


Fig. 5 An Acoustooptic Time-Integrating Correlator Using Hybrid Optical Waveguide Structure

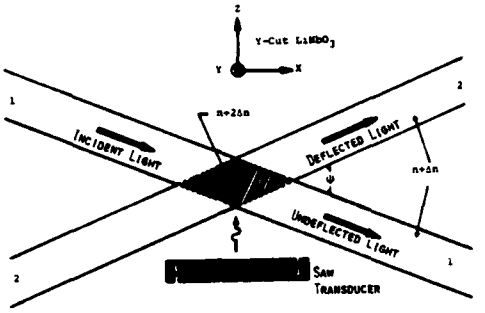


FIG.6 ACOUSTOOPTIC BRAGG DEFLECTION FROM SURFACE ACOUSTIC WAVE IN TWO CROSSED CHANNEL WAVEGUIDES.

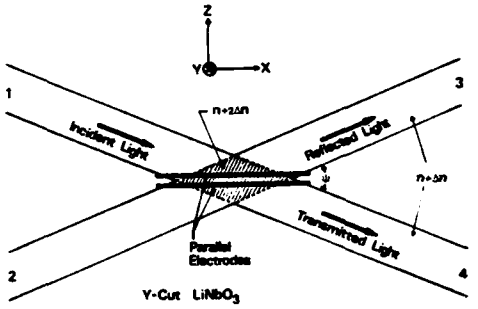


Fig. 7 Optical Channel Waveguide Double-Pull-Double-Throw Switch/Coupler Using Totle Internal Reflection

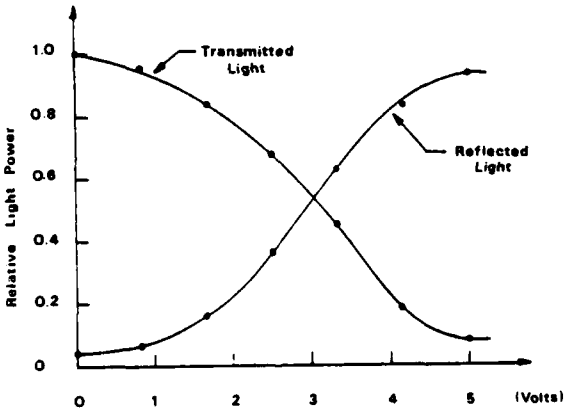


Fig. 8 Transmitted And Reflected Light Power Versus Drive Voltage

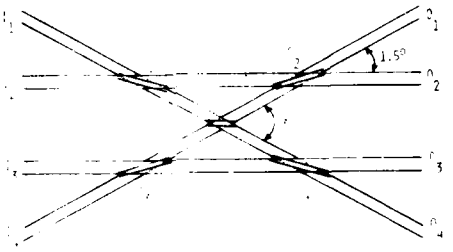
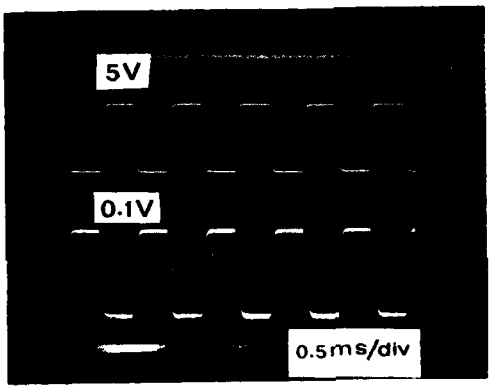


FIG.9 [illegible] INTERNAL REFLECTION.



(A) Modulating Signal Across S_1

(B) Output Signal From Output O_1

(C) Output Signal From Output O_2

(D) Output Signal From Output O_3

(E) Output Signal From Output O_4

Fig 10 Four Channel Outputs With I_2 Only Present, S_1 (Modulated), S_2 (ON), and S_5 (ON).

Discussion (Chen S. Tsai; Discussion Leader: Carl M. Verber)

Q. Is the ultimate limit on crosstalk in your switches just due to the quality of your waveguides?

A. We did some calculations based on discrete boundaries and the results showed that crosstalk should be smaller than what we measured. But we really have a diffuse boundary, shaped something like a Gaussian or complementary error function. Analytical calculations remain to be done to determine the optimal boundary shape. However, it may be that the best technique is not to try to get the best out of a single switch, but to just add an additional switch to better the crosstalk by a factor of 2 in db.

Q. It seems that with present technology we are very limited in how much we can change the refractive index, and this in turn seems to imply that we will always measure the size of integrated optics devices in maybe millimeters instead of microns. This same problem shows up with the electro-optic switches which have to come in almost at grazing incidence in order to work properly. Are we going to see improvements in this regard?

A. That's a very important question. With conventional titanium diffusion methods, developed at Bell Labs and other places, the most one can get is about 1% or 2%. But a recent development at Bell Labs uses an ion exchange technique to obtain an index change of 5% or 6%, and that's a very encouraging result. With an index change of about 1%, you get an angle between transmitted and reflected beams of around 12°, but with a 6% index change, the angle increases to almost 45%. This means the angles between switch paths can be increased, so that the aspect ratio of a switch array would be more like 1:1 rather than the present 20:1.

Q. You stated that it takes a change of voltage of 5 volts to activate the electro-optic switch. If the index change increases, does that imply that the switching voltage would be lowered?

A. No, the electro-optical effect isn't changing, so you still have to switch in and out of the critical angle.

C. Total internal reflection only works in principle for plane waves. As you have variation across the beam, you get leaky waves, and as you improve your angle to make the device smaller, the variation across the wavefront will increase. The larger angle will improve packing density of the switches, but leaky waves will be emphasized and thus crosstalk performance will suffer.

R. I think it is possible, but we haven't examined the problem. That's why I say we have only scratched the surface.

Q. In materials with a greater change of index, what is the ultimate smallest size for these switches?

A. This 4 x 4 switch is about 1.6 mm. from end to end, so the device itself is already very short. The thing which requires space is the channel separation in order to interface with fibers for input and output. This separation will be enhanced by using materials with a greater change of index. Bell Labs and Hughes built a 4 x 4 switch using directional couplers, but they had to use tapered guides, and that takes a lot of space. Here we are talking about a straight waveguide which takes much less space.

Q. Your switch and spectrum analyzer were on L_iNbO_3, yet the material GaAs can essentially incorporate all components, and so can Si, except possibly the laser. Could you comment on future directions here?

A. At this time only L_iNbO_3 has the best optical attenuation parameter, and light is very precious in a substrate. L_iNbO_3 is the best in terms of acousto-optic interaction efficiency, transducer efficiency, and optical attenuation. People have examined Si substrates with a special glass overlay, but it doesn't have nearly as good properties as L_iNbO_3. In time, maybe progress in different glasses and Si substrates will be made. With GaAs, the problem is to find a technique to generate surface acoustic waves, because efficiency is poor and acoustic attenuation is high. Bragg diffraction efficiency can be very good because the figure of merit is about 20 times better than for L_iNbO_3, but optical attenuation is high, about 4 db per cm. as opposed to .5 or 1 db for L_iNbO_3.

Q Could you trade off high figure of merit against coupling efficiency if you have your laser integrated right on the chip?

A. I haven't done that yet, maybe someone in industry has. Although L_iNbO_3 is presently the best material, that doesn't mean we don't do research in Si and GaAs. There are also other materials, such as arsenic trisulphide which has a figure of merit about 70 times greater than L_iNbO_3, whose optical loss is high and scattering is poor. But it is worthwhile to continue studies to see if one can get a better quality of film, and then we could utilize the higher figure of merit. What is the consequence of a higher figure of merit? We need lower acoustic power for efficient diffraction, and with lower acoustic power we have lower acoustic nonlinearity. Acoustic nonlinearity is proportional to acoustic power squared, so if you only need low acoustic power, perhaps arsenic trisulphide could give you high efficiency and potentially a higher dynamic range. Other materials, such as cadmium dioxide and tellurium, which have higher figures of merit than L_iNbO_3, are also worthwhile alternate substrates to study.

Optical Transfer Function (OTF) Synthesis Techniques For Noncoherent Processors

William W. Stoner

Science Applications, Inc., 3 Preston Court
Bedford, Massachusetts 01730

Abstract

Beginning in 1976, a new approach to optical data processing was introduced which retains many of the advantages while circumventing many of the disadvantages of conventional coherent optical processing. This approach exploits the Optical Transfer Function (OTF) of spatially incoherent imaging systems. It requires mastery of techniques for controlling OFTs; the field of OTF synthesis was born! We develop the subject by first motivating (Section 1) and explaining (Section 2) noncoherent optical processing, then delving into OTF sythesis (Section 3), and end by reviewing noncoherent optical processing applications and suggesting directions for further work in Section 4.

1. Introduction and Motivation For Noncoherent Optical Processing

Speed of light, diffraction limited parallel processing of large blocks of data - this is the potential of optical signal processing. Many years of effort to harness *coherent* optics have been, however, only a partial success; coherent optical processors remain vulnerable to parasitic coherent noise, fussy spatial filter alignments, problems with 2-dimensional spatial light modulators, etc. On the other hand, noncoherent optical devices (binoculars, slide projectors, eyeglasses, even precision microscopes and shaft encoders) are relatively carefree. This robust character suggests noncoherent optics would be superior to coherent optics for data processing. In the next section we shall trace light through both coherent and noncoherent processors, and illustrate the following specific advantages possessed by the noncoherent approach:

a) Noncoherent processors are insensitive to phase noise at the input, and consequently a wider variety of input devices (CRTs, self luminous objects, laser scanners) are accommodated. Furthermore, if a Spatial Light Modulator (SLM) intended for coherent optics is employed, phase defects in the SLM are inconsequential.[1]

b) Noncoherent processors possess an inherent multi-channel redundancy which provides a reduction in system noise due to scratches or dust.[2,3]

c) Larger dynamic ranges may be possible in coherent processors, thanks to the lower noise floor and the linearity of noncoherent processors with intensity which dovetails with the linear intensity response of many detectors.[4]

d) In some cases, noncoherent processors offer fewer critical mechanical and alignments than coherent processors.[3]

e) Noncoherent processing offers freedoms in filter (OTF/pupil mask) design which have no analog in coherent systems. This freedom permits pupil mask design choices and trade-offs between signal to noise and space-bandwidth product (Section 3).[5,6]

f) Noncoherent optics makes feasible larger space-bandwidth products (and processing gains) then coherent systems.[3]

2. General Principles

The idea of performing convolutions with a diffraction limited noncoherent optical system was pioneered by Lohmann[7] and Lowenthal.[8] A noncoherent imaging system naturally forms the convolution of the input image with the imaging system impulse response (point spread function). To convolve two functions, one must be implemented as the impulse response; the other is entered as an intensity distribution at the input plane, and the desired convolution is formed at the output plane (Figure 1). The point spread function (PSF) must be shaped into the desired convolving function by an appropriate choice of pupil plane mask. Let us stress that the pupil mask shapes the PSF by *diffraction* not by shadowcasting. This decision permits the processing of large arrays of data, but it also requires relatively narrowband illumination. Lohmann has shown that $(\lambda/\Delta\lambda)$ must be greater than the 1-dimensional space-bandwidth product of the system.[1]

A longstanding objection to this convolving technique has been the lack of "negative" intensity; most applications involve bipolar data or bipolar convolving functions. One answer to this objection is the encoding of polarity information with a temporal or spatial carrier frequency. The phase of a carrier need not be restricted to 0 or π, so complex values, as well as positive and negative values, may be represented.

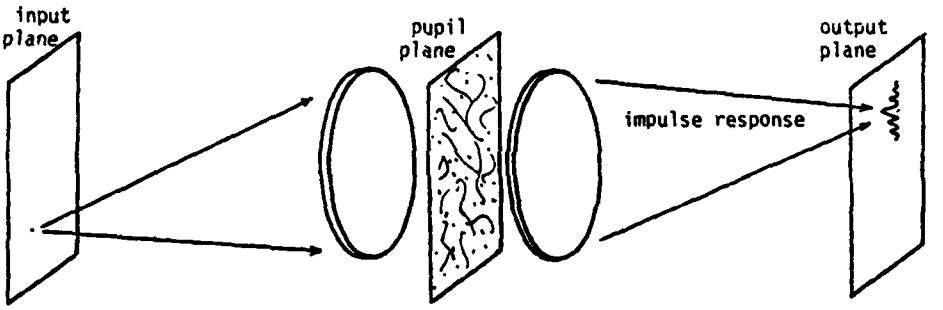


Fig. 1. Diffraction limited noncoherent convolver

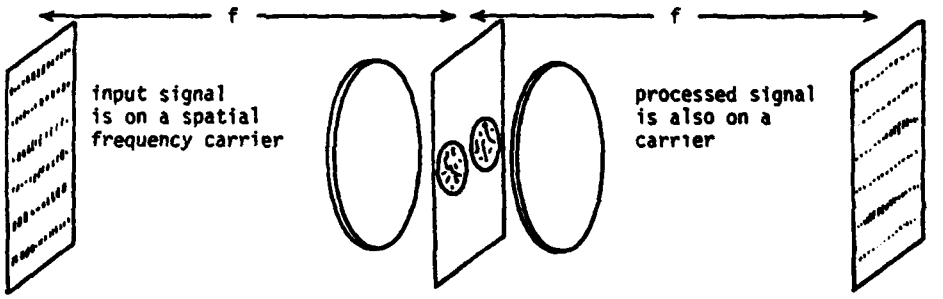


Fig. 2. Convolution of complex-valued data

2.1 Spatial carrier case

To avoid confusion between the spatial carrier case and the temporal carrier case, we shall cover each separately. A phase coded spatial carrier is generated on the PSF by limiting the transmitting area of the pupil mask to a pair of spatially offset sections (Figure 2). In essence, the resulting PSF is a Young's fringe pattern, modulated in intensity and encoded in phase through one's choice of pupil mask.

As is well-known, the PSF convolving technique is equivalently described as spatial filtering by the system OTF. The business portion of the OTF is frequency offset from DC (Figure 3), because the desired convolving function has been represented on the PSF by a modulated spatial carrier. In the output image, the processed data is also represented as modulation of a spatial carrier. Bandpass filtering and demodulation is needed to isolate and recover the processed data for display or further processing. Standard electronic filters and heterodyning mixers are suited for this step, if the output image is detected and scanned out in a sequential raster format.

In mathematical terms, the input output relationship of a noncoherent processor is given by[3]

$$O(x) = I(x) * F(x) , \tag{1}$$

where $I(x)$ represents the input intensity distribution, $F(x)$ represents the intensity PSF, and $O(x)$ represents the output intensity distribution. We have included only one spatial dimension for simplicity. The input illumination is assumed to be spatially incoherent and so each input point is imaged independently; and so the output intensity is the convolution (denoted by *) of the input intensity with the PSF. The equivalent frequency domain statement is

$$\tilde{O}(\mu) = \tilde{I}(\mu)\tilde{F}(\mu) , \tag{2}$$

where the tilde represents Fourier transformation. $\tilde{F}(\mu)$ is the OTF of the system. To understand how the OTF

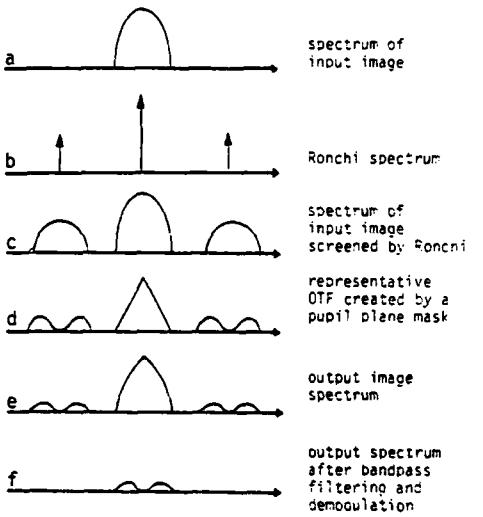


Fig. 3. Frequency domain viewpoint

is related to the pupil, we observe that the intensity PSF is given by the square magnitude of an amplitude $p(x)$:

$$F(x) = |p(x)|^2 \tag{3}$$

Fourier transforming this equation one finds

$$\tilde{F}(\mu) = \int_{-\infty}^{+\infty} \tilde{p}(\xi + \mu)\, \tilde{p}^*(\xi) d\xi \equiv \tilde{p}(\mu) \bigstar \tilde{p}(\mu), \tag{4}$$

where $p(x)$ and $\tilde{p}(\mu)$ form a Fourier transform pair, and where $\bigstar$ represents the correlation operation. From our knowledge of Fourier optics, we recognize that $\tilde{p}(\mu)$ is the pupil function that yields the amplitude $p(x)$ in the output plane. Thus the OTF is given by the autocorrelation of the pupil function. When the overall pupil consists of two spatially offset pupils $\tilde{p}_1(\mu)$ and $\tilde{p}_2(\mu)$

$$\tilde{p}(\mu) = \tilde{p}_1(\mu) + \tilde{p}_2(\mu - \Delta), \tag{5}$$

the OTF consists of an interaction part

$$(\tilde{p}_1(\mu) \bigstar \tilde{p}_2(\mu - \Delta) + \tilde{p}_2(\mu - \Delta) \bigstar \tilde{p}_1(\mu))$$

and a noninteraction part

$$(\tilde{p}_1(\mu) \bigstar \tilde{p}_1(\mu) + \tilde{p}_2(\mu) \bigstar \tilde{p}_2(\mu)).$$

The noninteraction part is simply the sum of the OTFs of each subpupil acting by itself, and this is highly constraining. This is made evident mathematically by their representation as autocorrelations. It should not come as a surprize that autocorrelation functions (e.g., $\tilde{p}_1(\mu) \bigstar \tilde{p}_1(\mu)$) are highly constrained, because they are the Fourier transforms of nonnegative real functions ($|p_1(x)|^2$). The interaction part of the OTF is represented by correlations, and is relatively unconstrained. Moreover, the spatial offset, Δ, between $\tilde{p}_1(\mu)$ and $\tilde{p}_2(\mu)$ translates into a spatial frequency offset which separates the interaction portion of the OTF from the noninteraction portion (Figures 3d and 5a).

2.2 Temporal carrier case

W. T. Rhodes[3] has shown how a Mach-Zehner interferometer may be used to optically superimpose two distinct pupils without a spatial offset (Figure 4). Once again the OTF consists of interaction and noninteraction terms, but in this case they are separated by inserting a phase shifting element in one of the interferometer paths. The overall pupil function becomes

$$\tilde{p}(\mu) = \tilde{p}_1(\mu) \exp(i\phi) + \tilde{p}_2(\mu). \tag{6}$$

The interaction portion of the OTF is given by

$$(\tilde{p}_1(\mu) \bigstar \tilde{p}_2(\mu) \exp(i\phi) + \tilde{p}_2(\mu) \bigstar \tilde{p}_1(\mu) \exp(-i\phi)),$$

and the noninteraction portion by

$$(\tilde{p}_1(\mu) \bigstar \tilde{p}_1(\mu) + \tilde{p}_2(\mu) \bigstar \tilde{p}_2(\mu)).$$

Notice that the noninteraction portion is insensitive to the phase shifting element. The output image may be recorded twice, once for $\phi = 0$ and once for $\phi = \pi$, and by subtraction, the effective OTF will be given purely by the unconstrained interaction portion of the OTF, $(\tilde{p}_1(\mu) \bigstar \tilde{p}_2(\mu) + \tilde{p}_2(\mu) \bigstar \tilde{p}_1(\mu))$. If $\phi = \frac{\pi}{2}$ and $\phi = -\frac{\pi}{2}$ is used, one obtains $\exp(i\frac{\pi}{2})(p_1(\mu) \bigstar p_2(\mu) - p_2(\mu) \bigstar p_1(\mu))$. These two independent interaction OTF are linear combinations of the positive and negative frequency portions of the interactive OTF obtained when a spatial offset is employed (Figure 3d and 5a). The analogy becomes complete if we shift the phase ϕ linearly in time; then the temporally offset interaction portions of the OTF become tagged with positive and negative carriers (Figure 5b).

The analysis following Eqs. (5) and (6) shows that the OTF synthesis problem is the same for both the spatial and the temporal carrier cases. One starts with an application that demands a particular filter function; then one searches for two pupils whose cross correlation yields the desired filter. This is a rather mathematical way to phrase the problem. Other approaches will be considered in the next section. Before moving ahead to OTF synthesis, however, we must make good the promise of Section 1 to illustrate how the advantages we claimed for noncoherent processors arise. We accomplish this by tracing coherent and noncoherent illumination through representative systems: Figure 6A: Collimated coherent light traverses the

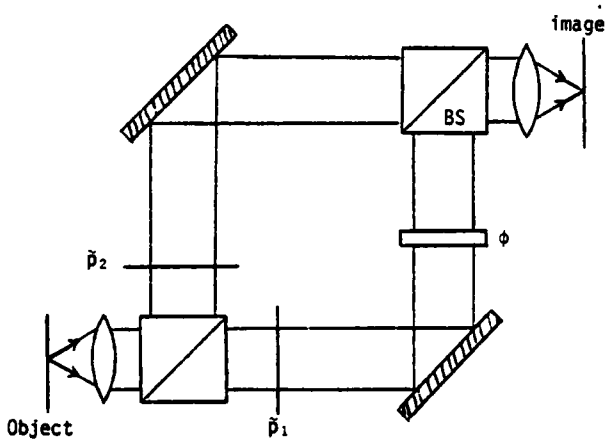


Fig. 4. Temporal offset noncoherent processor ($\tilde{p}_1$, $\tilde{p}_2$: pupil functions; BS: beamsplitter; ϕ: dynamic phase shifting element)

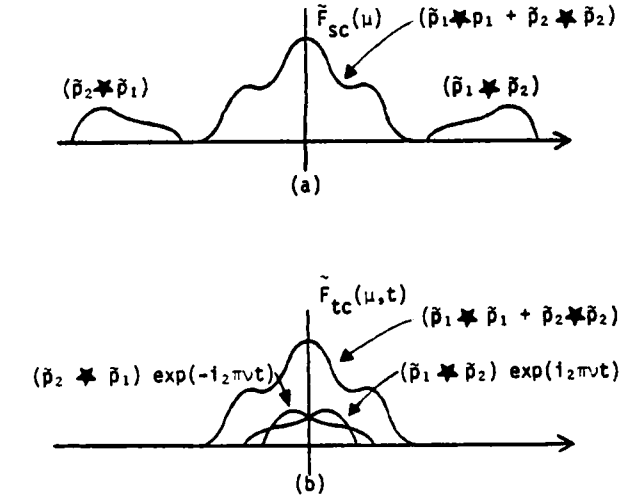


Fig. 5. OTFs obtained with carrier methods: (a) spatial carrier method, (b) temporal carrier method

input device, picking up noise introduced by thickness variations, striae or fingerprints. This phase noise propagates into amplitude and phase distribution downstream. Figure 6b: uncollimated, noncoherent light traverses the input device, its intensity unperturbed by optical phase noise. Since the noncoherent light is uncollimated, it beams out every element of the data pattern presented by the input device over the *entire* aperture of the optics downstrea, so that this information (represented by transmission variations across the input device) is carried redundantly to the output detector. Figure 6a: Because the input coherent light is collimated, dust on optics near the Fourier plane blots out tiny elements of the coherent transfer function, corrupting the intended function of the processor. Figure 6b: Dust on optics introduces only a slight degradation to the output image, because the noncoherent light is uncollimated and rays blocked by dust are compensated by rays carrying the same information which pass through other portions of the lens aperture. Figure 6a: Coherent light scattered by dust propagates to the output where it interferes with unscattered light, forming parasitic fringes that further degrade the image. The dynamic range of the output detector is also poorly utilized by the coherent processor, because the detector senses light intensity, while it is the light amplitude that carries the output information. As a consequence, an output dynamic range of only $\sqrt{D}$ is realized through the detector has a range of D before saturating. Figure 6b: noncoherent light scattered by dust forms a uniform output bias. Although noncoherent techniques generally introduce additional bias illumination, the resultant reduction in image contrast is acceptable; the output information is represented linearly by intensity, so that a bias level of B leaves a useful dynamic range (D-B) for the output information. It can happen that $(D-B) < \sqrt{D}$, and in this case coherent processing stands to provide the greater dynamic range. Even in here, however, the advantage may be negated by the inevitably greater noise floor of coherent optics. Figure 6a: The spatial filter must be precisely positioned because the coherent system uses collimated light. Figure 6b: In the noncoherent system, the transverse positioning of the pupil mask is irrelevant, except for extreme mispositionings which cause vignetting.

Although not reflected in the figure, there are two reasons why the noncoherent system may provide a larger space-bandwidth product than the coherent system: 1) The reduction in system noise afforded by noncoherent optics permits the use of multiple element lenses corrected over a wide field of view; coherent optical systems generally must accept simpler lenses with long focal lengths because multiple element optics tend to prohibitively increase system noise. The longer focal lengths overly constrain the achievable space-bandwidth product for a given overall processor size. 2) For a given lens aperture, the amplitude impulse response is wider and has higher sidelobes than the corresponding intensity impulse response. Consider for example $\sin(x)/x$ versus $(\sin(x)/x)^2$: full width at half maximum, 3.79 vs 2.79 radians; first sidelobe height, .21 vs .05. The sidelobes of the coherent system may be reduced by apodizing, but this will only serve to broaden the width of the central peak.

Of all these advantages, we believe the most important to be the superior noise performance of noncoherent optics. Although most of the evidence in this regard is experimental, a careful theoretical study by Chavel and Lowenthal concludes:[10]

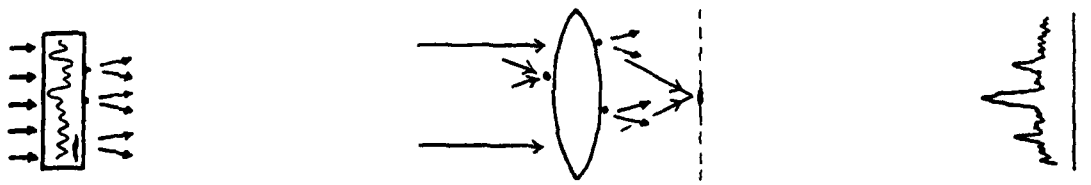


Fig. 6a. Representative coherent optical processor

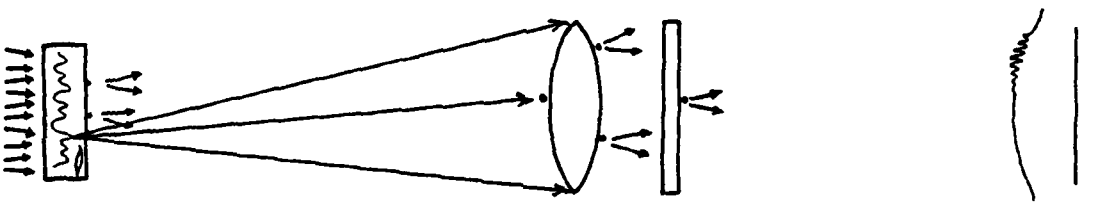


Fig. 6b. Representative noncoherent optical processor

" ... it is never a disadvantage, with respect to noise, to use incoherent illumination, and therefore that incoherent optical processing should be used whenever possible. In a coherent optical processing setup, clearness and optical quality of the components are of crucial importance everywhere, with the consequence that it is very difficult to overcome noise problems, whereas in an incoherent setup, noise due to the optical components is relatively harmless, and attention need only be given to amplitude defects of the object plane (grain noise, dust)."

3. OTF Synthesis Techniques

In every branch of signal processing, analog as well as digital, a considerable effort has been devoted to the design of filters for linear signal processing. In digital processing the effects of roundoff, sampling paramters, aliasing, interpolation (for conversion between polar and cartesian coordinates in sensor array processing, etc.), and the pros and cons of direct convolution versus frequency filtering using DFTs (discrete Fourier transforms) must all be carefully considered. Similar problems arise in analog signal processing, and in addition one must contend with device limitations such as charge transfer inefficiency in CCDs, sound wave attenuation, reflection and diffraction in SAWs, etc. Although work is still going on, progress in understanding and underlying physics of SAW devices has been sufficient to permit analytic design.[11] Our objective here is to provide the foundation for such design techniques in noncoherent pupil mask processors.

To avoid confusion we restrict the discussion of OTF synthesis to the spatial carrier case; as we have seen in Section 2, the close analogy between the spatial and temporal carrier cases permits results in one case to be useful in the other. One begins with a given convolution or spatial filtering requirement demanded by a particular application. One then follows the recipes of Lohmann[1] or Stoner[12] to install the desired frequency filter on a physically valid OFT. Equivalently, one may encode the desired convolution function with a spatial frequency carrier as a nonnegative, bandlimited and hence physically valid PSF. For our present purposes, it is convenient to work directly with such an encoded PSF. Recalling Eq. (3), we now ask what $p(x,y)$ corresponds to our $F(x,y)$. It is easy to solve for $|p|$ by taking the square root of F. It may also seem easy to multiply $|p|$ by an arbitrary phase factor, $\exp(i\phi(x,y))$, and to inverse transform to obtain $\tilde{p}(\mu,\nu)$. The catch is that $\tilde{p}(\mu,\nu)$ cannot go on forever, but must be zero for $\mu,\nu \gtrsim R/F\lambda$, where R is the radius of the processor lenses. Arbitrary choices of $\phi(x,y)$ lead to discontinuities in $p(x,y)$ which correspond to a $\tilde{p}(\mu,\nu)$ of infinite extent. If only the central region of this $\tilde{p}(\mu,\nu)$ is used, a smoothed approximation to $F(x,y)$ is obtained which will not exactly satisfy Eq. (3). It is clear that we need both a way of specifying good choices of $\phi(x,y)$ and a criterion for assessing the approximation we achieve to F and $\tilde{F}$.

The design techniques we require are quite closely related to the techniques developed to design computer generated holograms (CGH). The earliest CGSs were simply mimics of the holograms which can be obtained experimentally; in particular, they used a spatial carrier frequency to encode phase. An undesirable limitation of this early approach is that the spatial carrier uses up a large fraction of the plotting capacity of the plotter or scanner which is used to compose the hologram. Moreover, the diffraction efficiency of thin amplitude holograms is poor. Even earlier work of Tsujiuchi[13] demonstrated that true complex wave modulation was achievable by forming a sandwich filter consisting of an amplitude modulating transparency and a phase modulating transparancy. Registration of the transparencies is critical, however. Chu and Fienup[14] recently found a practical way to produce such phase and amplitude sandwiches. Multi-emulsion film, such as Kodak kodachrome color reversal film, contains film layers which can be separately exposed, by using exposures with different colors. After development, the resulting blue and red dye images can be used to control the amplitude and phase of a transmitted HeNe (red light) beam, because the red dye predominantly influences the phase of the transmitted beam, while the blue dye predominantly influences its amplitude. Because no spatial carrier is needed to encode phase, the image is produced on-axis and so this type of CGH contains no trace of analog to the reference beam of conventional holography. Hence the name ROACH --- Referenceless On-Axis Complex Hologram. Fienup's work[15] was directed towards using ROACHs in computer memories. This application demands good image fidelity, in order to achieve low bit error rates. For this reason careful work was done to improve the process of making good holograms. The photographic process had to be well understood, and beyond this, techniques were studied which attempt to find good choices for the phase, $\phi(x,y)$, of the holographically constructed image wavefront $p(x,y)$. Fienup's work followed up on an iterative phase construction technique developed by Gerchberg and Saxon.[16]

Many pupil functions for noncoherent optical processing have already been found without avail to an arsenal of phase construction techniques.[12] However, such an arsenal is extremely desirable, because noncoherent processing cannot be fully exploited without techniques for finding the optimal (in some sense to be determined) pupil function $\tilde{p}(\mu,\nu)$ for implementing a given $\tilde{F}(\mu,\nu)$ and $F(x,y)$. For this reason we took a broad look at the phase construction problem. It occurs in many disciplines, including electrical engineering, radio-astronomy, speckle interferometry, x-ray crystallography, elementray particle physics, and optics. We sought general solutions to the problem. The outline below gives an overview of the discussions that follow.

General Techniques For Phase Construction

- Given only $F(x)$ or $F(x,y)$:
  - 1-D case: analyticity and Hilbert transforms
  - our heuristic extension into 2-D
  - the feasibility of large-scale optimization
- Given $F(x,y)$ and a guess for $|\tilde{p}(\mu,\nu)|$:
  - Gerchberg-Saxon iteration
  - Dallas' algebraic solution

3.1 General techniques for phase construction

3.1.1 The use of Analyticity

This approach to the phase problem follows contributions made almost simultaneously be Walther and O'Neill,[17] Hofstetter,[19] and Goldberger, Lewis, and Watson.[19] It is interesting that these scientists are respectively from the disciplines of optics, electrical engineering, and physics. They followed a general approach first introduced by physicists studying scattering (Kramers and Kronig) and by electrical engineers studying electrical filters (H. W. Bode).[20] A serious limitation of this approach is that it covers only the 1-dimensional case, but we have developed a heuristic extension for the 2-dimensional case.

The analysis of the 1-dimensional case involves analytic continuation of the 1-dimensional PSF into the 2-dimensional complex plane; we shall recite the form of the analysis and refer the reader to the references[17-20] for their justification:

$$F(x) \longrightarrow F(z);\ F(z) = p(z)\ p^*(z^*) \qquad (7)$$

Analysis of $F(z)$ reveals that is is a product over its zeros in the complex phase. Furthermore, because $F(z)$ limits to $F(x)$ as z approaches the x axis, it follows that the zeros of $F(z)$ occur in complex conjugate pairs; if z_i is a zero, so is z_i^*. This ensures that $F(z)$ is nonnegative real on the x axis. The pairing of zeros introduces an ambiguity in $p(z)$, because there is no way to decide whether to pick out the factor $(z-z_i)$ or $(z-z_i^*)$ when constructing $p(z)$ from $F(z)$. Both choices lead to the same expression for $F(z)$, and consequently to the same expression for $F(x)$. The process of changing from $(z-z_i)$ to $(z-z_i^*)$ is called zero-flipping. Given $F(x)$, if we could only find one of the $p(z)$ we would have its zeros, z_i, and we could play the zero flipping trick to find any of the other $p(z)$ we wish. (Since the number of complex zeros may be zero, finite or infinite, there may be only one acceptable $p(x)$, a finite number, or an infinite number.)

An expression does exist for one of the $p(z)$. This special $p(z)$ has all of its zeros in the lower half plane. It is known as the minimum phase solution. This term comes from electrical circuit theory.[20] The

time delay of an electrical impulse through a linear circuit is related to the phase of the filter frequency response. Of all the electrical circuits with the same frequency response magnitude, the one which introduces the least delay has the least phase variation. In our context, an analogous statement holds. The minimum phase p(z) corresponds to the $\tilde{p}(\mu)$ which is maximally concentrated at μ=o in the following sense:

$$\int_0^a |\tilde{p}(\mu)|^2_{\text{min. phase}}\, d\mu > \int_0^a |\tilde{p}(\mu)|^2\, d\mu \quad \text{for all a.} \tag{8}$$

The phase, $\phi(x)$, of the minimum phase amplitude, $p(x) = |p(x)| \exp(i\phi(x))$, is given by the Hilbert transform of $\log(|F(x)|^{1/2})$:

$$\phi(x) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{\log(|F(y)|^{1/2})\, dy}{(y-x)} \tag{9}$$

We now have a recipe for computing the set of $\tilde{p}(\mu)$ which correspond to a given F(x): Use Eq. (9) to compute $p(x) = \sqrt{F(x)} \exp(i\phi(x))$. Fourier transform to obtain $\tilde{p}(\mu)$; then Laplace transform to obtain p(z), the analytic continuation of p(x). Compute the zeros of p(z) out to a large value of |z| (they all lie on the lower half plane). Obtain equally acceptable p(z) by zero-flipping. Fourier transform the corresponding p(x) to obtain new $\tilde{p}(\mu)$. Choose the $p(\mu)$ which is most easily fabricated.

3.1.2 Two dimensional extension

Given F(x,y), fix y to obtain the 1-dimensional function F(x,y'). Compute the zeros z_j of the corresponding analytically extended minimum phase amplitude, which we denote by p(z,y'). Consider $F(x,y'+\epsilon)$ for ϵ much less than the Nyquist sampling interval. By continuity, the zeros z_j of $p(z,y'+\epsilon)$ must approach the zeros of p(z,y') as ϵ approaches zero. If one stacks up the complex planes for increments of y', (Figure 7), the locus of the z_j are continuous curved lines occupying the Imag(z) < 0 half of the 3-dimensional space (z,y'). By continuity, these loci cannot end abruptly; however, they can move off towards infinite |z| . In this way it is possible for the character of the minimum phase p(z,y') to vary with y'.

We have not carefully investigated the properties of the function p(z,y'). It is clearly a close cousin to the minimum phase amplitude in the 1-dimensional case, and it limits to this amplitude if F(x,y) = F(x). Even if the Fourier transform of p(z=x,y) leads to an unsatisfactory $\tilde{p}(\mu,\nu)$, it can be expected to provide a good starting $|\tilde{p}(\mu,\nu)|$ for the iterative phase construction algorithm discussed in Section 3.1.4.

3.1.3 The feasibility of large-scale optimization

Given F(x,y), is it possible to find a least square fit F'(x,y) by optimizing parameters in the pupil? If F(x,y) has very little detail, and may be represented by a few samples, F(i,j), then $\tilde{p}(\mu,\nu)$ may also be adequately represented by a few samples $\tilde{p}(n,m)$. Optimization in ten variables can be done rather easily by computer, but the computation time for unstructured optimization increases as the square of the number of variables, and so the largest problems solvable with general purpose optimization routines like Fletcher-Powell on today's computers must have fewer than 200 variables. Unfortunately, optical processing has an advantage only when large amounts of data is to be processed, and therefore we are interested in F(x,y) which must be represented by millions of variables! Thus we would need to solve an optimization problem that is one hundred million times larger than the largest problems feasible. It is clear that brute force optimization procedures are useful only on small test problems that might be of interest in benchmarking other approaches; however, selective use of optimization in conjunction with other techniques remains a strong option.

3.1.4 Gerchberg-Saxon iteration

Given F(x,y) and a guess for $|\tilde{p}(\mu,\nu)|$, a technique originated by Gerchberg and Saxon[16] is applicable. A reasonable choice is $|\tilde{p}(\mu,\nu)| = 1$ within the open aperture of the imaging lenses, and $|\tilde{p}(\mu,\nu)| = 0$ outside, because this choice corresponds to the most efficient possible pupil function. This, of course, is no guarantee that the iterative algorithm will converge to a solution, because the given F(x,y) may be inconsistent with $|\tilde{p}(\mu,\nu)| = 1$.

The principle behind the Gerchberg-Saxon iteration is that the phase function over the pupil affects the output intensity distribution F(x,y). The given F(x,y) therefore contains implicit information about the phase $\theta(x,y)$ of $\tilde{p}(\mu,\nu)$. The problem is to combine this implicit phase information with $|\tilde{p}(\mu,\nu)|$. Gerchberg and Saxon proceed iteratively: Choose a random inital phase $\theta_1(\mu,\nu)$ for $\tilde{p}_1(\mu,\nu) = |\tilde{p}(\mu,\nu)| \exp(i\theta_1(x,y))$. Compute $p_1(x,y) = |p_1(x,y)| \exp(i\phi_1(x,y))$ by Fourier. The error in θ_1 causes an error in both $|p_1(x,y)|$ and $\phi_1(x,y)$. Correct $|p_1(x,y)|$ by replacing it with $\sqrt{F(x,y)}$. Compute a revised pupil phase $\theta_2(\mu,\nu)$ by Fourier transforming $\sqrt{F(x,y)} \exp(i\phi_1(x,y))$. This cycle from the pupil plane to the output plane and back has the effect of injecting into θ_2, θ_3, ..., θ_j, the phase information which is implicitly contained in F(x,y). Although the iteration may not converge to a pupil function which accurately produces an output intensity F(x,y), it can be shown that the mean-squared error between the estimates $F_j(x,y)$ and F(x,y) reduces (perhaps only infinitesimally) with each iteration. In the process, $\theta_j(\mu,\nu)$ and $\phi_j(x,j)$ also become more accurate.

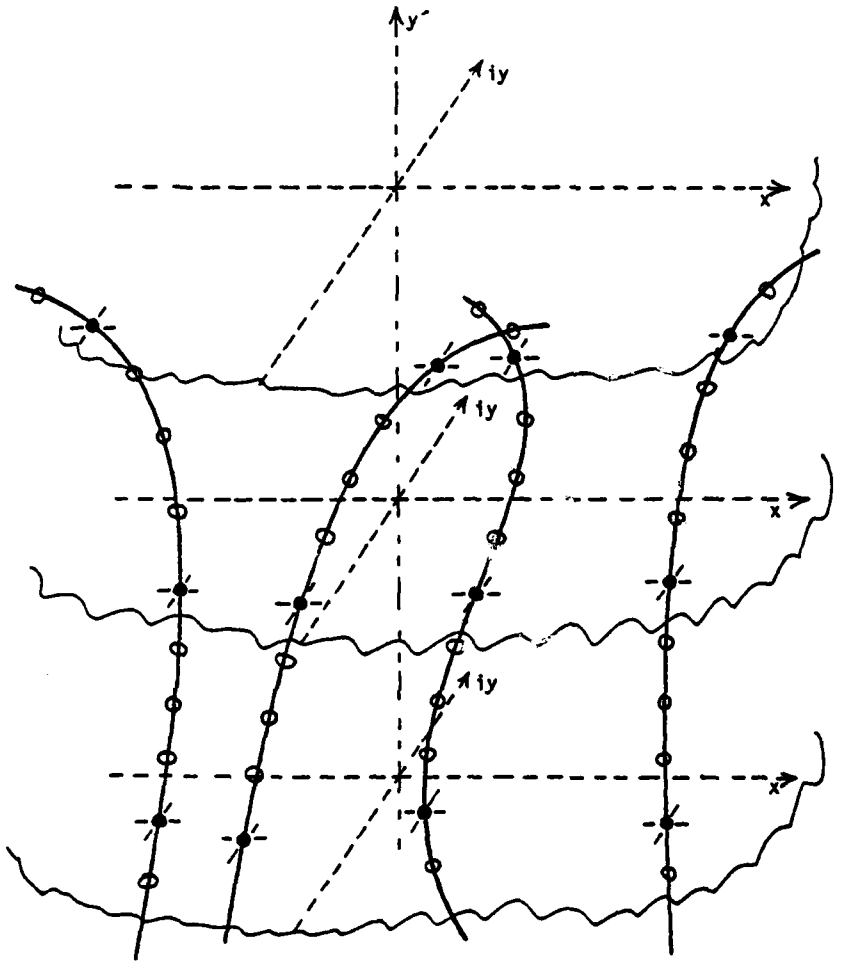


Fig. 7. Zeros of p(z,y')

Fienup has devised an over-relaxation technique which speeds up convergence, and has achieved some impressive results on related problems.

3.1.5 Dallas' algebraic approach

An approach which directly combines the information provided by $F(x,y)$ with the information provided by $|\tilde{p}(\mu,\nu)|$ was developed by Dallas.[21] The procedure is algebraic, and works with samples of $F(x,y)$ and $|\tilde{p}(\mu,\nu)|$. A criterion related to that of Nyquist is used to determine the sampling interval. First, all of the information is gathered into the pupil domain by applying a Discrete Fourier Transform (DFT) to $F(i,j)$. In analogy to Eq. (4), one obtains the autocorrelation of $\tilde{p}(m,n)$:

$$\tilde{F}(\ell,k) = \sum_{m,n} \tilde{p}(m+\ell,n+k)\tilde{p}^*(m,n) \qquad (10)$$

Together with the given values of $|\tilde{p}(m,n)|$, Eq. (10) may be progressively solved for $\tilde{p}(m,n)$. At each step, choices are encountered which may lead to alternate consistent solutions for $\tilde{p}(m,n)$. We view this as a strong point of the approach, because it leads to all possible solutions for the pupil function. The procedure has been carried out by hand on small 1-D test cases. However, we have found no reports of its successful application on large 2-D problems.

3.2 Pupil function replication in OTF synthesis

In the above discussions explicit reference has been made to the bandlimited properties of physically valid PSFs. This is perhaps overly confining. Lohmann[1], and Braunecker, Hauck, and Rhodes[6] advocate just the opposite! They point out that sampled convolution functions are perfectly acceptable in practical applications, and that this discrete structure of the PSF will tend to negate the problems associated with the choice of phase for $p(x,y)$. This being the case, they proceed to advocate a random phase, because this choice spreads the bandwidth, and leads to "homogenized" pupil functions. In other words, the choice of a random phase leads to greater redundancy in the pupil. There is a penalty imposed by this approach: the resultant PSF is faithful only on the discrete sampling lattice, and at points in-between the sampling lattice spurious garbage is introduced. Braunecker, Hauck, and Rhodes show that this garbage can be eliminated by simply replicating the pupil over a 2-D grid. In effect, they trade-off the bandwidth of the processor in order to reduce noise. We can intuitively understand their results by noting that the replicated pupil leads to a much finer sampling structure over the PSF than the basic unreplicated pupil achieves. This more precise sampling structure picks out faithfully constructed portions of the PSF and blanks out the garbage in-between. This effect is shown dramatically in Figure 5 of their paper. As a valuable side-benefit, the replication adds redundancy to the overall pupil, and this reduces the adverse effect of grain

in the pupil mask or dust on the processor optics. This approach is direct, relatively simple, and effective. It may be implemented with either an off-axis CGH or an on-axis ROACH. At this time, it is the best method known for OTF synthesis.

4. Applications and Directions For Futher Work

In the four years since OTF synthesis was first proposed as an alternative to coherent optical processing, over twenty papers have been published by seven groups worldwide. Reviewing this work we ask what the logical next steps are, and how this work might be channelled to achieve the greatest impact. In answering these questions, we are guided by the recent challenge of Guenther, Christensen, and Upatnieks,[22] who successfully demonstrated real time correlation tracking and target identification with a coherent optical processor anticipated to have an ultimate package size of 50 cubic centimeters. The proponents of noncoherent optical processing claim that it can do the same general complex valued or bipolar filtering and correlation operations as coherent optical processing, and that noncoherent processors can do so with superior performance. The application chosen by Guenther, Christensen, and Upatnieks is one that looks to be within reach of the available OTF synthesis techniques: CGHs of PSFs with relatively small space-bandwidth products are all that is needed for correlation of simple targets. Let's make themost of our presently developed noncoherent optics in this application! In this context, we must mention the run-off between coherent and noncoherent techniques for color coding of spatial frequencies recently performed by Indebetouw.[23] This result clearly demonstrated, at least in this application, the superiority of noncoherent over coherent processing.

In general, more work needs to be done on real time noncoherent processing, in order to move into specific applications. The work of Gorlitz and Lanzl[24] with a color encoded pupil in conjunction with a color video camera, and the work of Barrett, Greivenkamp, Gordan, Gmitro, Chui, and Swindell[25] on real time noncoherent optical processing for transaxial tomography are examples of significant progress in this direction.

At the present time, the most accessible applications are those which permit a fixed pupil mask to be used. Dynamically changing the pupil mask will require further work on pupil mask design and OTF synthesis, as well as work on programmable spatial light modulators. Examples of applications using fixed masks include 1) detection of a small class of targets or signals, 2) reformatting of data, 3) Fourier transformation using the chirp transform algorithm[12], 4) frequency excision of fixed frequency bands, and 5) image processing such as in transaxial tomography or image characterization.[23]

Applications which appear less accessible at this time include correlation or convolution of two signals both having large space-bandwidth products, because this will involve improvements in techniques for designing pupil functions and constructing pupil masks. Work should be done to push and define the limits of the presently known techniques. Work should also be done to explore the feasibility of using present day acousto-optical devices and SLMs for dynamically controlling the pupil plane.*

5. References

1. A.W. Lohmann, Applied Optics 16, 261 (1977).
2. G.L. Rogers, Opt. Laser Technol. 7, 153 (1975).
3. W.T. Rhodes, "Noncoherent Optical Processing with Two-Pupil Hybrid Systems," presented at the International Optical Computing Conference 23-26 August 1977 (SPIE 118, 1977)
4. W.W. Stoner, "System Roles for Noncoherent Optical Processing," (SPIE 209, 1979).
5. A.W. Lohmann and W.T. Rhodes, Applied Optics 17, 1141 (1978).
6. B. Braunecker, R. Hauck, and W.T. Rhodes, Applied Optics 18, 45 (1979).
7. A.W. Lohmann, Applied Optics 7, 561 (1968).
8. S. Lowenthal and A. Werts, C.R. Acad. Sci. Ser. B 266, 542 (1968).
9. The following treatment and notation borrows freely from Reference 5.
10. P. Chavel and S. Lowenthal, JOSA 68, 721 (1978).
11. R.H. Tancrell, *IEEE Trans.* SU-21 1, (January 1974).
12. W. Stoner, Applied Optics 17, 2454 (1978)
13. J. Tsujiuchi, Progress In Optics 2, 131 (North-Holland, 1963).
14. D.C. Chu, J.R. Fienup, and J.W. Goodman, Applied Optics, 12, 1386 (1973)
15. J.R. Fienup, "Improved Synthesis and Computational Methods for Computer Generated Holograms," Thesis, Stanford University, 1975.
16. R.W. Gerchberg and W.O. Saxon, Optic 35, 237 (1972).
17. A. Walther, Optics Acta 10, 41 (1962).
18. E.M. Hofstetter, IEEE Trans. IT-9, 119 (1963).
19. M. Goldberger, H. Lewis, and K. Watson, Phys. Rev. 132, 2764 (1963).
20. H.W. Bode, Network Analysis and Feedbak Amplifier Design, Chapter XIV (Robert E. Krieger, Huntington, NY, 1975).
21. W.J. Dallas, Optik 44, 1, 45 (1975).
22. B.D. Guenther, C.R. Christensen, and J. Upatnieks, IEEE JQE, QE-15, No. 12, 1348 (December 1979).
23. G. Indebetouw, Applied Optics 18, 4206 (1979).
24. D. Gorlitz and F. Lanzl, Optics Commun. 28, No. 3, 283 (1979).
25. H.H. Barett, J. Greivenkamp, S.K. Gordon, A.F. Gmitro, M.Y. Chiu, and W. Swindell, Optics Commun. 28, No. 3, 287 (1979).

* Partial support for this work has been provided by RADC/ET, Hanscom AFB under contract F19628-78-C-0117.

Discussion (William W. Stoner; Discussion Leader: William T. Rhodes)

C. In looking for future directions, it is clear that real-time operations should be investigated. I've always claimed that coherent optical techniques aren't very well suited to the real world and that they are going to remain in a laboratory environment. Non-coherent systems are very robust, and it has only been a question of how you can harness such systems to do optical signal processing. The particular task of tracking and correlating a target against a background has been done successfully with a coherent system, and that should be within reach of noncoherent optics.

C. What we need are run-offs between coherent and noncoherent systems. Recently an article described a runoff between a coherent and noncoherent processor in an application for image characterization. The object was to take an image and pick out the different spatial frequency bands by color coding. They found that, in this particular case anyway, the noncoherent techniques were superior. Runoffs are something that could be done in universities very easily, a good thesis project. Another thing that would be interesting to look at with a spatially incoherent system is the dollar bill recognition experiment.

Q. I was visiting a company and saw a machine they built which was basically a slide viewer. They used an incandescent source with a small filament, and they did spatial filtering to enhance the contrast of the slides. Do you foresee that these techniques would allow spatial filtering to be done in normal optical instruments, such as binoculars? Or a camera that has a knob you can turn to sharpen the image?

A. It's a simple enough operation, yes. There was a company selling a pupil mask which was just an apodized mask that you put over your lens. It was a photographically made disk that was opaque in the center and gradually got more transparent. If you are doing standard darkroom work and you want to blur out a picture, you might do that by de-focusing. But when you put one of these masks on, the mis-focused OTF improves. So there are applications for pupil masks that have been around for awhile. If you want to sharpen the image, you are in trouble because you need an OTF that inhibits low spatial frequencies relative to high, and that violates the positivity requirement.

Q. Does it require some electronic post-filtering?

A. You need that if you want to do anything like edge enhancement. But there are other techniques, like Gordon Rogers' procedure of using a large bias term.

C. There is still the broadband nature of a scene you photograph with a camera. A limited number of OTF syntheses will work with broadband illumination, but if you are trying to do a variety of information processing tasks that would be of interest to this community, then you need to stick with narrowband illumination.

R. Whenever you want a sharp feature in the Fourier domain, you need narrowband illumination. But if you have just broad features, then you can loosen up the tolerance on the spectral purity of the light. Now let's look at the filter I was talking about, such as a notch filter for an extreme example. The OTF is designed for one wavelength, and if you use another wavelength everything gets scaled. If the notch is really narrow, then the two notches won't be on top of each other and the filter will be fuzzed out.

C. Maybe one could set up an optical system so that part of the OTF synthesis was done by the front end. The optical system or lens might be apportioned somehow so as to have an OTF that emphasized whatever it was you were looking for. One can imagine that being useful in IR search and track where you are trying to extend the range at which you pick up the target. A question arises concerning the S/N ratio, because your choice is to block off part of the aperture and do your processing with a reduced amount of light flux. If the choice is to have a large aperture, which is very heavy and may be impossible to get into orbit, versus having a couple small apertures that you can move around, then you can synthesize a pupil and get the effect of a large aperture.

C. In coded aperture imaging you use an annulus and convolve the annulus with itself. It occurs to me that you could use some other function, like the 1/0 filter. There are probably a number of common aspects between coded aperture imaging and OTF synthesis.

C. You might even be able to tailor the optics in the front end with whatever is convenient to make in terms of the detector array.

Parallel Incoherent Optical Matrix-Vector Multipliers

Joseph W. Goodman, Antonio R. Dias, Kristina M. Johnson and David Peri

Department of Electrical Engineering, Stanford University
Stanford, California 94305

Abstract

A review of various approaches to constructing incoherent optical matrix-vector multipliers is presented. Progress in the experimental realization of such processors is outlined and analytic studies of performance limitations are summarized.

Introduction

Since the invention of the laser, much work has been devoted to the realization of coherent optical information processing systems. The most attractive attribute of such systems has been the inherent parallelism with which they can perform data processing operations. It appears to be generally recognized that the chief limitations to the speed at which data processing operations can be performed by such systems lie with the input and output devices. While the optical systems themselves are capable of performing highly parallel operations, nonetheless, the data is generally input through a single serial electronic channel and output in a similar fashion.

While the data to be processed may originate in a serial fashion in many data processing problems, nonetheless, there clearly exist problems in which the data is more fundamentally parallel in nature. Examples include all those situations in which parallel arrays of sensors gather acoustic or electromagnetic information. In such cases it seems unfortunate to be forced to multiplex the information into a single serial data stream for entry into a parallel optical data processor.

In an attempt to find ways to overcome the input/output bottleneck, we have been exploring the properties of a certain class of optical processors that are capable of parallel input and output of data. Much of the stimulation that led to these ideas was provided by the work of Bromley and his associates,[1] who devised a serial incoherent optical matrix-vector multiplier. Our work differs from this earlier work primarily through the emphasis we have placed on parallel input and output, a requirement that radically changes the architecture of the processor.

In the material to follow, we first introduce the reader to the basic concepts underlying the parallel matrix-vector multiplier. Next, we discuss several different optical realizations of such a processor. Some of the theoretical limitations to system performance are then briefly discussed. Finally, the current status of the experimental systems that have been built is reviewed.

Basic System Configuration[2,3,4]

By way of background, and to establish notation, we briefly review the nature of the data-processing operation we are attempting to perform. Consider a column vector $\vec{x}$ with components $x_0, x_1, \ldots, x_{N-1}$ representing N data samples to be processed. Our goal is to multiply this vector times an $N \times M$ stored matrix $\underline{H}$, having element h_{mn} in the n^{th} column and m^{th} row. The result of this operation is a vector $\vec{y}$ with components $y_0, y_1, \ldots, y_{N-1}$ representing the processed data. Thus,

$$\vec{y} = \underline{H}\vec{x} . \tag{1}$$

A single element y_m of the output vector is related to the elements of the input vector and the stored matrix by the simple equation

$$y_m = \sum_{n=0}^{N-1} h_{mn} x_n \tag{2}$$

Note that computation of a single y_m requires, in general, the performance of N multiplications and $N-1$ additions. We shall describe methods for performing these operations optically and in parallel for all M components of the vector $\vec{y}$.

We remark in passing that the data processing operation expressed by Eq.(1) is a very common one in practice. With suitable attention to sampling rates, any linear data processing operation (space-invariant or space-variant) can be reduced to this form.

At this point it is convenient to assume temporarily that the elements of $\vec{x}$, $\underline{H}$, and $\vec{y}$ are all non-negative and real, thus allowing them to be represented physically by light intensities (for $\vec{x}$ and $\vec{y}$) or

intensity transmittances (for $\underline{H}$). Methods for extending system operation to bipolar-real or complex data are covered in the section to follow.

With reference to Fig. 1, let the elements of the input vector $\vec{x}$ be entered in parallel as brightness or radiance values on an N-element array of mutually incoherent sources. These sources could be light-emitting diodes, or they could be lasing diodes.

The function of the first box labeled "optics" in Fig. 1 is to spread the light from each source into a vertical column of light incident on a unique column of the matrix mask. Thus this portion of the optical system must effectively image in the horizontal direction but spread light uniformly in the vertical direction.

The matrix mask can be realized physically as a transparency consisting of $N \times M$ partially transparent cells. The intensity transmittance of a given cell must be proportional to the value h_{mn} of a particular matrix element, the proportionality constant being the same for all cells and assuring that all elements lie between 0 and 1. The transparency of a cell can be controlled by either adjusting its optical density or opening a transparent subcell with area proportional to the matrix value of interest. Of course, a combination of these techniques could also be used. If the matrix $\underline{H}$ is fixed, the transparency can be photographic. If it is desired to change the matrix rapidly (e.g., at television frame rates), a controllable transparency, such as a liquid crystal light valve, can be used.

As the light passes through the matrix mask, $N \times M$ analog multiplications are performed optically. The function of the second box labeled "optics" is to perform the analog additions required to obtain the output components. Thus the second set of optics produces, on each of the M photodetectors, light intensities proportional to the sum of light intensities transmitted by individual rows of the matrix mask. In order to maintain the integrity of the M output channels, these optical components must image in the vertical direction while effectively spreading the light from each mask element horizontally.

Each output detector element generates a current that is proportional to a different element y_m of the output vector. The M output components are available in parallel.

Achieving Bipolar and Complex Operations[3,5]

In practical signal processing problems, it seldom suffices to accept only non-negative and real inputs and to subject them to only non-negative and real matrix multiplications. Rather, it is necessary to extend the domain to at least bipolar real vectors and matrices, and preferably to complex-valued vectors and matrices.

There exist many methods for representing complex numbers by real and non-negative components. For example, we could use four non-negative and real numbers representing basis components separated by 90° in the complex plane, or three non-negative and real numbers representing components separated by 120°. In both cases, to arrive at a unique decomposition, one must adopt some conventions, such as two out of four (in the four component decomposition) or one out of three (in the three-component decomposition) or the components must always be zero. Such encoding circuits entail the use of non-linear devices (half-wave linear rectifiers), which broaden the bandwidth of the driving circuits. More importantly, each of these methods uses more than two input devices (LED's) to represent each complex number.

A more efficient method for performing complex operations can be realized with a two-component decomposition. If it were possible to realize bipolar, real input and mask values, then the decomposition

$$\begin{bmatrix} \vec{y}_R \\ \vec{y}_I \end{bmatrix} = \begin{bmatrix} \underline{H}_R & -\underline{H}_I \\ \underline{H}_I & \underline{H}_R \end{bmatrix} \begin{bmatrix} \vec{x}_R \\ \vec{x}_I \end{bmatrix} \tag{3}$$

would be possible, where subscripts R and I indicate bipolar real and imaginary parts. However, in practice, we cannot directly input bipolar vector elements nor multiply by bipolar matrix elements, since the system is incoherent. Nonetheless, we can bias the mask and input vector to guarantee non-negative elements throughout, in which case the components of the output vector obtained are given by

$$\begin{bmatrix} \vec{y}'_R \\ \vec{y}'_I \end{bmatrix} = \begin{bmatrix} \underline{A}_R + \underline{H}_R & \underline{A}_I - \underline{H}_I \\ \underline{A}_I + \underline{H}_I & \underline{A}_R + \underline{H}_R \end{bmatrix} \begin{bmatrix} \vec{b}_R + \vec{x}_R \\ \vec{b}_I + \vec{x}_I \end{bmatrix} \tag{4}$$

where $\vec{b}_R$ and $\vec{b}_I$ are bias vectors,

$$\begin{aligned} \vec{b}_R &= \left(b_0^R, b_1^R, \ldots, b_{N-1}^R\right)^t \\ \vec{b}_I &= \left(b_0^I, b_1^I, \ldots, b_{N-1}^I\right)^t \end{aligned} \tag{5}$$

,

while A_R and A_I are bias matrices,

$$\underline{A}_R = \begin{bmatrix} a^R_{00} & a^R_{01} \cdots a^R_{0,N-1} \\ a^R_{10} & \cdots\cdots \\ \vdots & \\ a^R_{M-1,0} & \cdots\cdots a^R_{M-1,N-1} \end{bmatrix}, \quad \underline{A}_I = \begin{bmatrix} a^I_{00} & a^I_{01} \cdots a^I_{0,N-1} \\ a^I_{10} & \cdots\cdots \\ \vdots & \\ a^I_{M-1,0} & \cdots\cdots a^I_{M-1,N-1} \end{bmatrix}. \tag{6}$$

After detecting the elements $\vec{y}'_R$ and $\vec{y}'_I$ at the output of the matrix-vector multiplier, the task remains of removing the unwanted components introduced by the biases and producing bipolar real and imaginary parts. $\vec{y}_R$ and $\vec{y}_I$, electronically. The desired vectors are recoverable from the measured vectors through the relations

$$\begin{aligned} \vec{y}_R &= \vec{y}'_R - \left(\underline{A}_R\vec{b}_R + \underline{A}_I\vec{b}_I\right) - \left(\underline{H}_R\vec{b}_R - \underline{H}_I\vec{b}_I\right) - \left(\underline{A}_R\vec{x}_R + \underline{A}_I\vec{x}_I\right) \\ \vec{y}_I &= \vec{y}'_I - \left(\underline{A}_I\vec{b}_R + \underline{A}_R\vec{b}_I\right) - \left(\underline{H}_I\vec{b}_R + \underline{H}_R\vec{b}_I\right) - \left(\underline{A}_I\vec{x}_R + \underline{A}_R\vec{x}_I\right). \end{aligned} \tag{7a}$$

The terms $\vec{y}'_R$ and $\vec{y}'_I$ represent measured data, and are therefore known. The first parentheses in both (7a) and (7b) consist of bias terms, known a priori, and therefore removable after detection. The second set of parentheses in both equations consists of the product of the known bipolar matrices with the known bias vectors, and therefore can be removed after detection.

The last parentheses in the equations contain products of the known bias matrices with the unknown bipolar input vectors, and therefore are unknown a priori. However, if the bias matrix $\underline{A}_R$ has all column vectors equal to a single column vector $\vec{a}_R$, and $\underline{A}_I$ has all columns equal to a single column vector $\vec{a}_I$, then these particular bias terms can be removed by adding an M+1st constant row to the stored matrix and an M+1st detector to the output array, and subtracting from each measured output component the properly weighted output of the M+1st detector.

We conclude that it is possible to multiply an N-element complex input vector times an N×M stored complex matrix by inputting data on a 2N-element array of incoherent sources, using 2N×(2M+1) cell transmission mask, and detecting intensities on a 2M+1 detector array. The use of biased real and imaginary parts yields the minimum number of input and output elements.

System Realization Using Discrete Optical Components

The first system that was built, and which indeed is still under construction, is a system that uses discrete optical components (spherical and cylindrical lenses) to perform the desired operations. Figure 2 shows the geometry of the optical system. Lens L_1 is cylindrical and collimates the light that is spreading in the direction normal to the LED array. Lens combinations L_2 and L_3 consist of spherical elements and serve to image each LED as a vertical line of light incident on the matrix mask. Lens L_4 is a field lens that brings the central rays transmitted by each cell of the mask to a common vertical column centered on the vertical detector array. Lens L_5 is cylindrical and images each row of the matrix mask onto a separate detector element. All elements are off the shelf items, and the total cost of the optics was about $200.

The LED array consists of 10 Texas Instruments TIES-36 infrared diodes, each capable of emitting a milliwatt of 9100Å radiation into a cone with a half-power beam width of 25 degrees (±12.5 degrees). The LED's are mounted with 3.5mm center to center spacings, resulting in an overall array length of 44mm. The ten separate drivers for these sources serve to convert an applied input voltage into a current passing through the corresponding LED. These drivers were designed in such a way that the slopes and biases of their characteristic curves can be controlled by externally applied voltages, such as might be provided by a microcomputer, for example. The inclusion of this extra flexibility led to a considerable price in terms of the linearity of the input-output characteristics of the combined driver and LED. Figure 3 shows a typical plot output-light-power vs. applied voltage for such LED and driver combination. The curve represents the output power as a function of voltage applied to the drivers. Each channel is operated with a bias, which places the excursions in the most linear region of the curve.

The masks used in this system have been constructed in binary form, with each matrix element represented by a transparent cell on an opaque backgtound. Initially, the mask is constructed such that the area of each transparent opening is proportional to the desired value of that particular matrix element.

The particular matrix-vector operation for which the mask has been designed is the discrete Fourier transform (DFT). For such an operation, the number of output elements equals the number of input elements (N=M), and the real and imaginary parts of the ideal matrix are

$$\underline{H}_R = \left[\cos \frac{2\pi nm}{N}\right] , \quad \underline{H}_I = \left[\sin \frac{2\pi nm}{N}\right] . \tag{8}$$

When biases are added, the actual forms of the stored matrices are

$$\underline{H}'_R = \frac{1}{2}\left[1 + \cos \frac{2\pi nm}{N}\right] , \quad \underline{H}'_I = \frac{1}{2}\left[1 + \sin \frac{2\pi nm}{N}\right] . \tag{9}$$

In practice we have found it helpful to view the matrix-vector multiplier as a kind of optical memory, for which a single stored analog value (i.e., a single matrix element) can be read out by turning on a particular LED source and a particular detector element. We invariably find that, once an ideal mask has been made, an optical read-out of the stored weights shows that there exist substantial departures from the ideal values. There are many sources of error, including vignetting present in the optics and errors inherent in the way a height or width modulation of an opening in the mask affects the stored weight read out optically. We have therefore found it generally necessary to construct the mask through an iterative process, in which we begin with an ideal mask, read out the optically stored weights, and then construct a second mask with elements modified in such a way as to bring the optically stored weights closer to their ideal values. The important point here is that one should not think of the matrix as being represented solely by the stored mask, but rather by the entire optical system, from the LED array to the detector array.

Figure 4 illustrates two masks, the first being ideal, and the second a corrected version that yields more accurate optically stored weights.

The detector array chosen for this system is a Centronics LD20-3, a 20 element PIN diode array with all elements accessible in parallel. The detectors have peak sensitivity at about 9000Å. Only 11 of these elements are used by the system, 10 for the output components and one for the bias channel. The length of the array is 20mm. Each detector element is followed by a low-noise pre-amplifier and amplifier mounted in the system. The gains and biases of all detector channels can be controlled by externally supplied electronic signals. The output bandwidth of each channel is approximately 10 MHz. The noise-equivalent-input-power (NEP) of the detector-amplifier combinations is a few times 10^{-12} watts/$\sqrt{\text{Hz}}$.

Figure 5 shows a typical output sequence read out by means of a Reticon CCD array (not the Centronics array). This particular sequence represents the real part of the transform of the input sequence (1010000000).

The system we have described here should be fully operational by the end of the summer of 1980. Corrections of limited complexity can be made to the electronically read weights by microcomputer control of the gains and biases of all input and output channels. Figure 6 shows a photograph of the system (excluding power supplies and microcomputer). The LED array and drivers are located at the far left, while the detectors and amplifiers are located at the far right. The entire length of the system is slightly more than one meter.

System Realization Using Multimode Planar Waveguides

A second realization of the matrix-vector multiplier is illustrated in Fig. 7. In this case the discrete optical elements of the previous system are replaced by two arrays of planar multimode optical waveguides. As will be seen, this realization is far more compact than that described earlier.

Each LED source now emits light into a planar multimode optical waveguide. With a proper choice of dimensions for the guide, multiple internal reflections of the light within the guide generate a relatively uniform column of light at its output. Figure 8 illustrates the measured distribution of light intensity for a guide with dimensions 76.2mm×22mm×1mm. Small irregularities in this distribution are smoothed by the finite size of the mask openings. In addition, iteration of the matrix mask can compensate for any remaining non-uniformities.

Light leaving the matrix mask passes into a similar set of guides rotated 90° with respect to the previous set. The optical power from each opening in the mask is spread to cover the entire end of the guide covering its respective row. The detector element intercepts a portion of the light at the guide output.

Figure 9 shows a photograph of the optical components of this system. The entire length of the optics is approximately 15 cm.

The weights stored in such a system when an ideal DFT mask is used have been read. The results show a total rms error of 5.3%. By iteration of the mask, it should be possible to reduce these errors to about 1%.

A new set of LED drivers and detectors is under construction for this system. The linearity of the new electronics is much improved. Figure 10 shows the measured characteristic of a single input channel.

The detector array used in this system will again be a Centronics LD20-3. The overall optical efficiency of this system is about a factor of 7 better than that for the system described earlier. Again the design bandwidth for each channel is 10 MHz.

System Realization Using Optical Fibers

A third realization of this type of system has been constructed with optical fibers. As illustrated in Figure 11, each LED fires into a bundle of optical fibers. In the system that has been constructed, there are 20 input ports, each consisting of 200 fibers. Each fiber bundle is split into sub-bundles, with the number of fibers in one sub-bundle being proportional to the weight of one of the elements in the stored matrix. At a single output port, a set of sub-bundles from all input ports is gathered to form a single output bundle. The light from that bundle falls on an output detector.

The matrix elements are stored in the fiber wiring arrangement of this system. Each weight is quantized to a number of levels equal to the number of fibers in a single input bundle. Figure 12 shows a photograph of the optical system that has been built; this system is capable of performing a 20-point DFT.

The chief advantage of this particular realization is its very high optical efficiency, which in turn allows high throughput speed (due to improved S/N ratio). The chief disadvantages are two-fold. First, the matrix cannot be changed or iterated, since it is inherent in the wiring diagram of the optical fibers. Second, and more important, the accuracy of the system is limited by the uniformity with which light can be coupled into all of the fibers. Polishing of the ends of the fiber bundles is a critical step in achieving uniformity.

In the particular system that was built, the rms error of the stored weights was measured to be 8.5%. We believe that this number can be significantly improved by better polishing of the ends of the fiber bundles.

Analysis of Performance Limits

An extensive analysis of the performance limits of incoherent optical matrix-vector multipliers has been performed and will be published as a technical report in the near future. Here we can only sketch the approaches taken and the results obtained.

The nonlinearities of the systems considered reside primarily in the LED drivers. Both the LED's themselves and the PIN photodiodes are extremely linear. Measured LED driver characteristics have been modeled with a polynomial fit, and a theoretical analysis of the output errors for both deterministic and stochastic inputs has been performed.

The nonlinear distortions present in the output are, of course, a function of how hard the input electronics are driven. The level of the input excursions influences the levels of the measured output excursions, which, in turn, influence the signal-to-noise ratio achieved at the output. The chief source of noise, in this regard, is thermal noise originating in the detector pre-amplifiers. Thus, there is a direct tradeoff between linearity and signal-to-noise ratio at the output.

Our analysis indicates that the planar waveguide realization is capable of achieving a S/N of 32.3 db at a level of total harmonic distortion (power) (THD) of 0.06% (32.2 db). With the use of more powerful (commercially available) LED's, we project that a S/N better than 52 db can be achieved with a similar level of THD. Of course, S/N and THD can be traded off against one another to some extent, if desired.

A second serious limitation to system accuracy arises from errors in the optical weights achieved to represent the matrix. In this regard, it is possible to define a merit factor (MF) associated with the mask as

$$MF = 10 \log_{10} \frac{\text{Total av. ideal output power}}{\text{Total ave. error power}} . \qquad (10)$$

For the "ideal" mask used initially in the planar waveguide realization, the value of the merit factor is 26 db. We anticipate that iteration of the mask will increase the value of the merit factor to about 40 db.

We should emphasize that the anticipated performance levels in our system are not necessarily the ultimate limit. An industrial organization with greater resources in terms of money and equipment could no doubt achieve performance specifications significantly better than we are quoting here.

Applications

The problems to which a matrix-vector multiplier such as we have described might be applied are strongly influenced by the number of channels and per-channel bandwidths that can ultimately be realized. Optical components capable of 150 MHz bandwidth are available on an off the shelf basis. The use of lasing diodes with even GHz bandwidths can be envisioned, but at a substantial increase in cost. The complexity of the operations that can be performed is limited primarily by the number of parallel input channels that one can build.

A processor having 100 channels, each with 100 MHz of bandwidth, would be capable of a throughput rate of 5×10^9 complex data samples per second. Such a system might find useful application in parallel beamforming for antenna arrays, high speed pattern classification by the nearest neighbor rule, or in the performance of very rapid coordinate transformations.

We would mention that the rapidly evolving optical communication technology will have a direct impact on cost and performance capabilities. In addition, the development of very high speed integrated circuits will surely impact our ability to build large numbers of high-speed input and output channels. Lastly, we mention that this type of system can be combined with CCD devices or SAW devices in a corner-turning architecture, as shown in Figure 13, to perform two-dimensional discrete Fourier transforms, or equivalently, products of input matrices (entered column by column) with stored separable 2-D matrix transformations.

Acknowledgment

The work reported here has been supported by the Air Force Office of Scientific Research and by the National Science Foundation.

References

1. M.A. Monahan, K. Bromley, and R.P. Bocker, Proc. IEEE 65, 121 (1977).

2. J.W. Goodman, A.R. Dias, and L.M. Woody, Optics Letters 2, 1-3 (1978).

3. J.W. Goodman, A.R. Dias, L.M. Woody, and J. Erickson, "Application of optical communication technology to optical information processing," LASL Conference on Optics '79, Proc. SPIE, 190, 485-496 (1979).

4. J.W. Goodman, A.R. Dias, L.M. Woody, and J. Erickson, Proc. ICO-11, Madrid, September 1978.

5. J.W. Goodman and L.M. Woody, Appl. Opt. 16, 2611-2612 (1977).

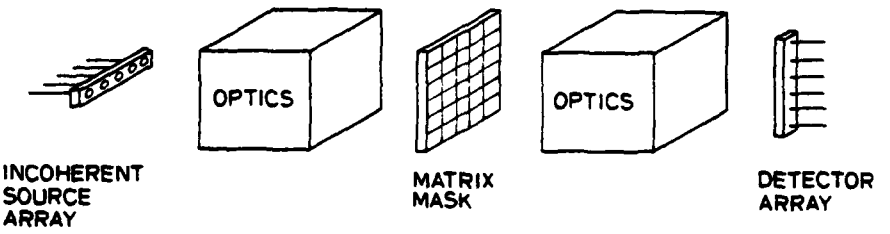


Figure 1. Basic Configuration of Incoherent Matrix-Vector Multiplier.

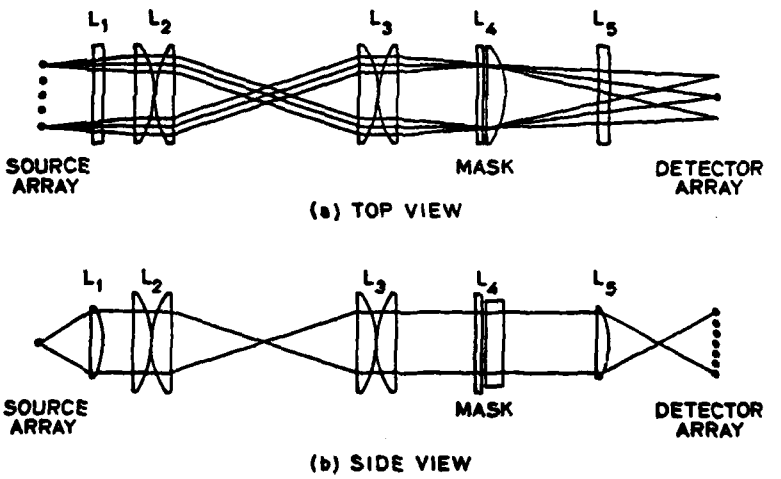


Figure 2. Optical System for Performing Matrix-Vector Products.

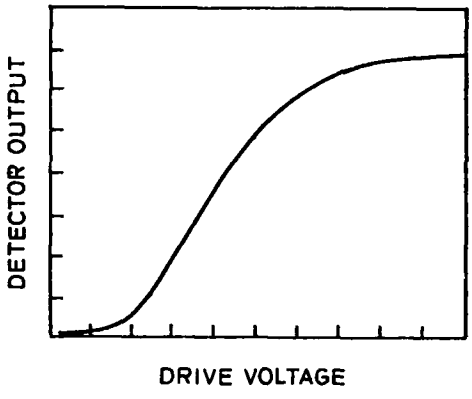


Figure 3. Light Output vs. Applied Voltage for First-Generation LED Array.

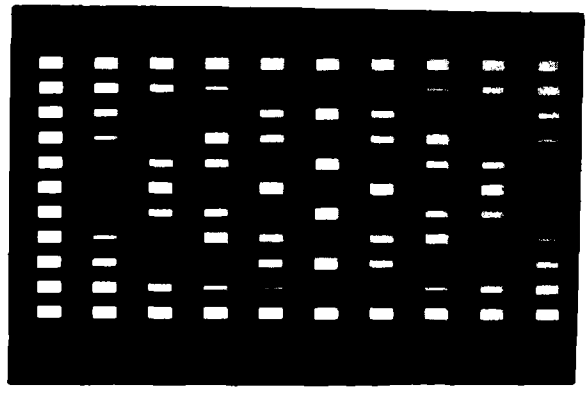


Figure 4a. Binary Masks for One Part of the DFT Ideal One Iteration.

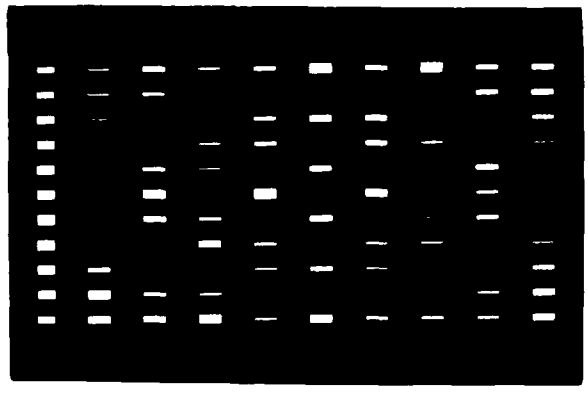


Figure 4b. Binary Masks for One Part of the DFT After One Iteration.

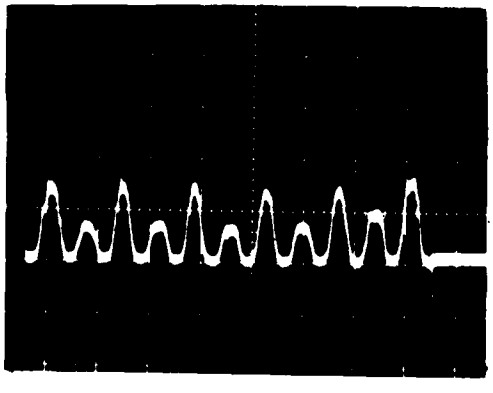


Figure 5. Typical DFT Output.

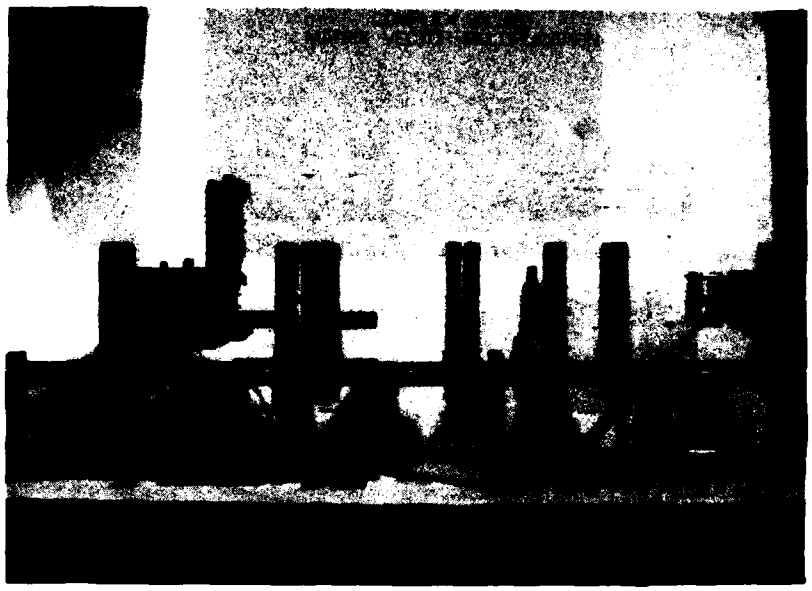


Figure 6. Photograph of the Overall System.

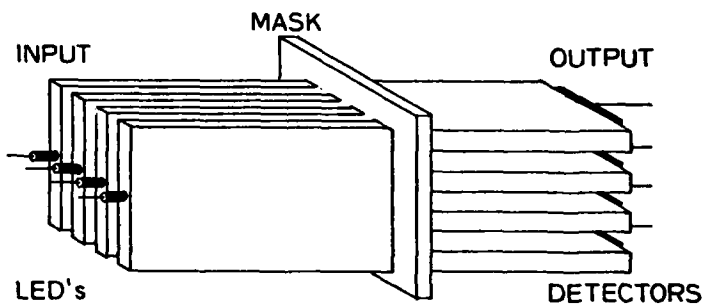


Figure 7. Configuration of System Based on Planar Waveguides.

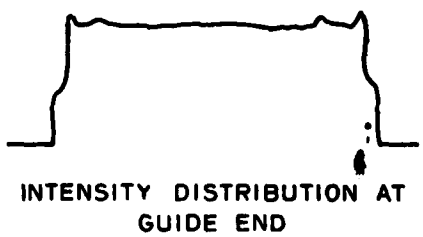


Figure 8. Intensity Distribution Measured at the End of a Single Guide.

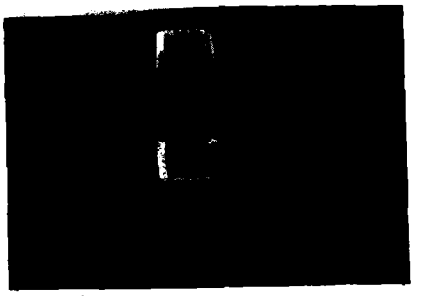


Figure 9. Photograph of Planar Waveguide Optics.

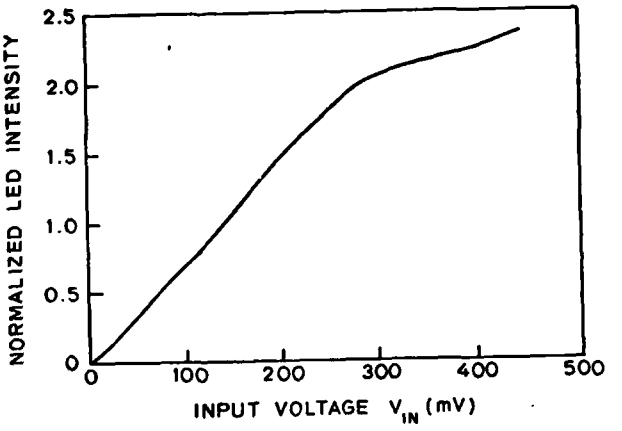


Figure 10. Measured Input-Output Characteristic of Second-Generation LED Sources.

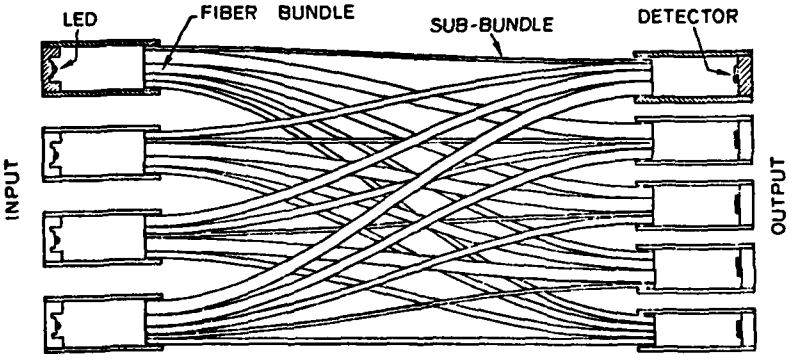


Figure 11. Configuration of Fiber-Optic Matrix-Vector Multiplier.

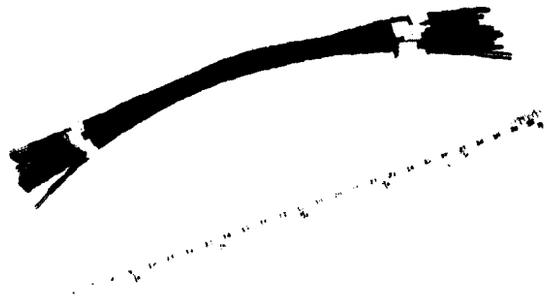


Figure 12. Photograph of Fiber Optic Realization.

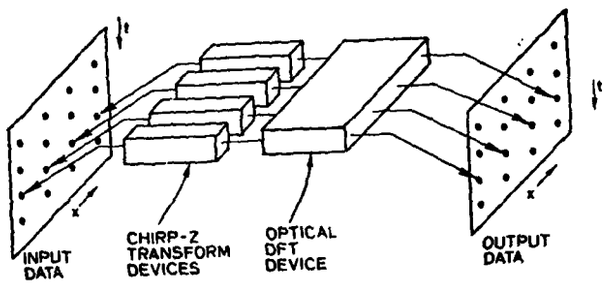


Figure 13. Combined SAW-Incoherent Optical System for Performing 2-D Fourier Transforms.

Discussion (Joseph W. Goodman; Discussion Leader: Thomas K. Gaylord)

Q. You pointed out the equivalence between matrix-vector multiplication and 1-D space variant operations. Of course it is the difference between discrete and incoherent, typically, and continuous and coherent. In terms of John Walkup's talk, do you see these two different ways as competing or complementary?

A. They're complementary in the sense that we first conceived of this system after thinking about space variant systems much as what John described. I think there is a fundamental difference between the two in that here we are trying to emphasize a massive parallelism in input and output and get away from serial channels. In more conventional coherent optical systems you have serial channels, for example acousto-optic inputs. Thus I see the approaches as complementary, not competing.

Q. It seems you have a trade-off, you have to produce n x n sources. Will the price always be high?

A. For n x n operations, I need n sources. There are two aspects to the price question. One is the electronics and the other is the sources themselves. These sources have a 10 MHz bandwidth, 1 mw per source, and cost about $20 for quantities of one. If you want to go to higher power, 10 to 20 mw. LED's, the price escalates quite a bit, you'll pay about 10 times as much per source. The electronics associated with a channel in the 10 MHz regime is not really difficult. I know one could build a 100 MHz per channel system, but I don't know what the cost would be. As to the future, someone made the statement earlier that whatever happens to coherent light valves depends on what happens in display projection. That is, display projection is the financially attractive application of these devices that motivates people to develop them. I would respond in the same way that fiber optics communications systems is a big and growing business, and I'm fairly optimistic that with time there will be a downward trend in the cost of these sources and an upward trend in their capabilities. Already the Japanese have reported the integration of a number of avalanche photodiodes on a single substrate.

C. Maybe a figure-of-merit is something like dollars per space-bandwidth product.

R. I think one needs dynamic range in there too, to make it a little more complicated. It is a difficult question to arrive at a figure-of-merit.

Q. In your future directions you mentioned the idea of generating masks by E-beam techniques. Do you mean photographic techniques or using an E-beam spatial light modulator?

A. I was thinking of E-beam systems that write directly on photographic film in nanoseconds with micron accuracy. I think one could take advantage of that technology if one wants to control the weights very accurately in a repeatable way. But you open a whole new aspect of this which I forgot to mention. You could conceivably make the matrix mask a light valve and change it dynamically in time, which might make for some very interesting applications.

Q. Is there any application for which you need real time and update in real time capability?

A. I think so, probably in the area of adaptive arrays, you can do beam forming with a system like this.

C. You mentioned that one version of the system has the input amplitude modulated on a carrier to reduce distortion. I wonder if you could eliminate the two channels and get around the positive-negative problems if you modulated in amplitude and phase, assuming you could build a complex mask.

R. I haven't been able to find a way to do that, and the reason is that I can't modulate the mask. If you use a modulation with the real and imaginary parts of the signal representing the real and imaginary vector components, at the end you have to be able to extract the inner product of the real part with its appropriate part of the mask and the inner product of the imaginary part with its appropriate part of the mask. I see no way to sort those out unless one somehow tags them as they go through the mask. By amplitude modulating the signal you gain dynamic range, but by introducing more complex modulation I don't see that you buy anything more unless you can modulate the mask.

Q. How large an array can you realistically imagine? You had 10 elements here.

A. I feel secure in saying that one could build a 100 element array. Someone told me he has a 1024 element array, but it doesn't have a 10 MHz bandwidth per channel. I don't think the device technology of making large arrays is the real thing that frightens me about a large number of channels, it is the parallel electronics.

C. You might use polarized light to get your bipolar range, where you have two polarizations and one acts as a reference phase for the other.

R. That might work, we really haven't thought much about using polarizers to do that.

Q. Earlier Dr. Lee showed us a device which was completely planar and consisted of sources, detectors, and a signal processor. My question is, which is best, a 3-D processor or a 2-D processor? You gain in simplicity with 2-D but you pay a price in throughput rate since you must wait for two waves to propagate past one

another. In essence, you are putting one dimension of the weights into time and the other dimension of the weights is fixed. I find it very appealing to make a 3-D system and not be constrained to a plane.

A. I concur with your analysis, and I think there's applications for both of them, depending on the system. In the 2-D case, you get a very nice match between CCD's at the input to parallel channels, the SAW device, and the A-O interaction. You can get several orders of magnitude increase in throughput by making the chirp Z-transform SAW's and then having this 3-D parallel device. It is a question of whether you have an application that needs 2 or 3 orders of magnitude increased throughput. I think they are complementary.

C. If you are looking for applications, people have talked about imaging Fourier transform spectrometers. If you had a spectrometer with 1000 x 1000 spatial elements, and if you wanted to look at the spectrum at those million points 10 times a second, then you are about to get up into this type of throughput. It is hard to think up a technique that would be relatively broadband and extract that much information from a scene by any other method than an imaging Fourier transform spectrometer, because filters tend to have variable fields of view and may not have the spectral bandwidth that you would like.

Digital Optics

Alan Huang
Stanford Electronics Laboratory
101B Durand Building
Stanford, California 94305

Introduction

The purpose of this paper is to try and predict or suggest the future direction of digital optics. In order to accomplish this the state-of-the-art will be briefly reviewed. The advantages and disadvantages of optics will then be discussed in an abstract manner. Some possible applications matched with these capabilities will then be examined. An example of a simple optical digital computer will then be presented to demonstrate some of the techniques, problems, and possiblities of using digital optics.

The State-of-the-Art of Digital Optics

Current efforts in the area of digital optics are centered around optical memories, logic devices, and methods of performing arithmetic.

Optical Memories

There has been a resurgence of interest in read only and read-write memories based on video disk technology. The effort has been led by some large companies such as IBM, CDC, and Honeywell. The projected application for such memories involve magnetic tape replacements and large data base storage.

Optical Logic

One approach towards optical logic involves laser diodes [BAS]. Unfortunately, it was difficult to demonstrate any advantage of this technology over comparable electronic approaches [LAN]. This early work is being closely reconsidered by the integrated optics community. The push by optical communications has prompted the development of several devices capable of performing digital operations. Controllable waveguide couplers are one such device. They have been used for logic and simple A/D units by Taylor [TAY]. Another such device is the integrated optic versions of Fabry-Perot interferometers which are fast and require very little energy to latch [GAR]. While many novel devices have been demonstrated integrated optics is still haunted by the capabilities of electronic GaAs devices. The viability of integrated optics will depend heavily on whether it can ultilize some of advantage of optics other than just bandwidth.

The incremental pumping of a laser cavity can also be used to perform logic. IBM has some of the early patents in this area. The problem with laser logic involves the energy, hardware, and cycle time required by this approach. A new approach using dyes has been suggested by Johnson [JOH]. A similar approach is currently being pursued as a means of achieving sub-picosecond switching times.

Logic operation can also be performed by interfering the outputs of a properly programmed holographic associative memory. This holographic approach to logic is being pursued by Guest and Gaylord (Georgia Tech) [GUE]. This approach is capable of performing simple Boolean operations by table lookup. It can be extended to perform simple combinatoric logic in a manner similar to a programable logic array. This can be taken one step further. The basic table lookup mechanism can also perform some types of arithmetic with little more complexity than the orginal Boolean operations. Similar work is mentioned by Mnatsakanyan [MNA].

Another approach towards digital logic relies on controllable transparencies. This approach involves mating integrated circuits technology and optics. The basic mechanism involves using an array of photoconductors each modulating the transparency of a separate liquid crystal cell. Such work is being done by Lee (U.C.S.D) [LEE] and Orlov [ORL]. The object is to produce arrays capable of performing logic operations. A first binary image would activate the photoconductor and thus control the liquid crystal cell. These cells would then interact with a second binary image.

Spatial modulators such as the Hughes liquid crystal light valve and the ITEK prom can also perform logic operations in an array like manner. Collins (Ohio State) has used a light valve with feedback to implement optical flip-flops and other logic operations [COL2]. Strand (U.S.C.) has used a light valve in a variable grating mode to achieve A/D conversion along with various logic functions [STR]. Hudgins (ITEK) [HIS] and Cindrich (ERIM) have respectively mentioned that similar operations are possible with the ITEK prom and photo-elastomers. What is so intriging about these approaches is that they achieve logic operations in bulk. Thousands of operations are just as complex as a single operation. One drawback is that these mechanisms are relatively slow. However, it must be remembered that these devices were orginally developed for dynamic range rather than speed. It seems reasonable to assume that faster non-linear mechanisms are possible.

Logic arrays seem to be a common goal for researchers involved with holographic logic, controlled transparencies, and spatial light modulators. One reason for this is that it ultilizes the parallelism of optics. The question of how to use these devices still remains. A clue to this can be found in the TSE computer suggested by Schaefer [SCH]. This is a two dimensional computer intended for processing images. While the implementation is a little awkward the basic idea of processing images with Boolean logic arrays is quite interesting.

Optical Arithmetic

It is also possible to perform arithmetic directly and bypass the necessity of implementing Boolean operators. The idea was initally suggested by Basov [BAS]. This led to the work by Huang [HUA1]. This work has since been continued by Huang [HUA2], Collins [COL1], Stoner [STO], and Cindrich [CIN]. Collins impressed the cyclic nature of the residue number system on the cyclic nature of polarization. Cindrich proposes impressing the residue number system on the spatial domain using integrated optical waveguide couplers.

Another approach towards performing arithmetic directly involves the use of optical convolution to perform multiplication. This has been suggested by Whitehouse (NEL) and Casasent (Carnegie-Mellon) [CAS].

The Advantages of Optics

Massive Parallelism

Lenses, mirrors, prisms, etc. can easily handle, with their inate parallelism, millions of resolvable points at the same time. The difficulty is how to use this parallelism. Part of the problem is conceptual. We think sequentially and thus it is hard to apply this parallelism. Another part of the problem is that such a system must be almost completely optical since any electrical interface will place a practical limit on the amount of parallelism which can be achieved. A hundred optical channels implies a hundred sources and detectors.

Non-Planar Propagation

Optical signals can pass through each other without interference. This is unique. Electrical signals have to be guided by wires and guarded from interacting with other electrical signals. The problem is: why use optics when wires are so cheap? Advances in computers are quickly changing this. Speed requirements force interconnections to be short (less than 6 inches) and of very large bandwidths (10 ghz). A large portion of a high speed machine such as the CRAY-1 or AMDAHL 470 is devoted to coaxial cables and terminating resistors. The main problem is now one of communications rather than switching speed. This provides a reason for using optics. The question remains how? If each signal requires a source and detector, the electronics overhead would limit the practicality of any such approach.

Disadvantages of Optics

Detection and Restimulation

It is very difficult to get one optical signal to affect another optical signal. Typically to achieve this, the optical signals must be detected and the result restimulated. This involves time and hardware. The electronics required makes this process highly questionable. If a pn junction is used to detect the light and a pn junction is used to generate the light, why use light in the first place.

The time it takes for detection and restimulation is especially crucial in a conventional sequential computer. New pipelined architectures have changed this. The detection and resimulation overhead can be absorbed by the pipeline. The quest is now for throughput rather than just speed. The question becomes how many detections and restimulations can be accomplished in parallel?

Optical Logic

Many schemes have been explored to implement optical logic. The difficulties are in speed and interfacing. Pipelining has altered some of the constraints on optical logic. Parallelism can be used to compensate for speed. The interfacing problem is a restatement of the detection and restimulation problem. The output of most optical logic has to be detected and then restimulated to produce another optical signal. This increases the complexity and limits the speed. The interfacing must be minimized or else the required electronics will squelch the parallelism which can be achieved by the optics.

Appropriate Technology

When is it proper to use electronics and when optics?

There is a reciprocal relationship between optics and electronics which results from the very nature of electrons and photons. It is very easy for electrons to affect other electrons. As a result, the interaction of electronic signals is easy but they are awkward for communications since they have to be guided by wires and shielded from other electrical signals.

On the other hand, it is very difficult for a photon to affect other photons. As a result, optical signals are good for communications but awkward to interact with.

Some other guidelines are that optics should not try to mimic electronics directly. The contest quickly becomes economic. As an example, logic using integrated optics might be feasible but the electronic version costs only five cents. In order to compete, optics must find some way to ultilize either its parallelism or its non-planar propagation ability. This is very difficult. Such a system must be almost purely optical since any hybrid approach will suffer from the limitations of an electronics interface.

Why Digital Optics?

Now that we have some idea of the state-of-the-art of digital optics and some of the inherent advantages and disadvantages of optics we are faced with the question of what to do. Who needs the capabilities of optics?

One possible application is computers. The reason for this can best be seen by studying some of the problem which limit the development of future computers. Some of these problems are clock skew, bandwidth, interconnections, and the Von Neumann bottleneck. What these problems have in common is communications. The current difficulty is no longer one of switching speed but rather how to get the right data to the right place at the right time.

Clock Skew

The clock skew problem can be seen in Figure 1. This problem occurs when the switching time approaches the propagation delay. An error will result unless the different input signals arrive at the gate simultaneously. This implies that the path lengths of all the inputs have to be almost identical. This is particularly difficult for a planar interconnection technology. These difficulties place restrictions on the geometry of the processor. The CRAY-1 computer is a good example of this problem.

Bandwidth

Another hardware problem is one of bandwidth. The logic which is used today triggers on either the rising or falling edge of a signal. See Figure 2. To keep these edges sharp requires the preservation of the higher harmonics of the signals. The propagation of a 1 nanosecond pulse requires a transmission line with a bandwidth much greater than 1 ghz. This implies that most of the off chip interconnections have to be accomplished with terminated coaxial lines. A large postion of the volume of current processors such as the AMDAHL 470 are already occuppied in this manner.

Interconnections

Another area of difficult is in the architecture of computers. Processors have traditionally been designed with centeralized control. As the processor grow larger and faster the flow of information becomes congested at the centralized control point. In order to relieve this problem the control is being distributed. The more it is distributed the more coordination is needed. This requires more communication paths. The number of interconnections grows very quickly. N centers requires N(N-1) bilateral communications paths. See Figure 3. The typical approach around this problem is to share a high speed bus. This trades time for hardware interconnections. Unfortunately, there is a limit to this time multiplexing. This limits the number and speed of such centers. The alternative is to increase the number of interconnections but this adds to the complexity and control overhead. This inability to communicate limits the complexity which can be achieved.

The Von Neumann Bottleneck

An even more fundamental problem is the Von Neumann bottleneck [BAC]. The structure of a conventional computer is shown in Figure 4. The difficulty is that all the instructions and data have to travel between the CPU and memory. This is agravated by the fact that only one item can be handled at a time and that an address is needed for each item. Currently the majority of the time required to perform a particular computation is involved in figuring out where to get and put the data. This bookkeeping overhead and sequentality are some of the fundamental limitations of throughput. Why this problem exists and how it can be avoided can best be understood by reviewing some computer fundamentals.

A classical finite state machine [TOR], as shown in Figure 5, is the ancestor of all present day computers. It consists of latches connected to a combinatoric logic unit. The latches preserve the outputs of the combinatoric circuits for one cycle such that they can be used as inputs on the next cycle. This circuit is completely parallel and has no Von Neumann bottleneck. There is no worry about where to get and put the information since the interconnections between the latches and logic are fixed. The system is also parallel since all the latches can be accessed independently at the same time.

The trouble begins when more state variables are needed. The number of interconnections between the latches and logic quickly increase. A binary encoding technique is used to reduce N interconnections to $\lceil \log_2 N \rceil$ interconnections. This is the source of the Von Neumann bottleneck.

A modified finite state machine is shown ir Figure 6. The number of interconnections is greatly reduced. Unfortunately, the characteristics of the system have also been altered. The system is now sequential rather than parallel. The addressing mechanism can only access one latch at a time. It must also know which latch to get from or put to. Furthermore, the characteristics of the latches themselves have been altered. Since they are addressed in a random manner the latches must now hold their contents for an indefinate amount of time. This is in contrast with the classical finite state machine (Figure 5) in which the latches need only hold the information for 1 cycle.

The bottleneck exists because in electronics, it is more practical to have $\lceil \log_2 N \rceil$ wires rather than N wires. This compromise and the resulting bottleneck is also reflected in the the address decoder of every memory chip as well as the protocols of each bus. In order to remove the Von Neumann bottleneck from the hardware a practical means of communicating N channels in parallel must be found. Given the right conditions, this is quite simple for optics.

Optical Memories

Optical computers have been suggested before. One of the major drawbacks was the development of an optical memory. The main difficulties are memory addressing and storage material. Unfortunately, optics tried to implement the the memory needed by a modified finite state machine as shown in Figure 6 rather than the simpler memory needed by the classical finite state machine as shown in Figure 5. The memory of the modified finite state machine has the compromises required by an electronics technology, the Von Neumann bottleneck. The potiental parallelism of optics was thus restricted from the onset. The addressing mechanism was slow, awkward, and expensive. Beam deflectors, page composers, and detector arrays were needed. The storage material was required to store the information indefinitely rather than for just one cycle as in the classical finite state machine. The search for such a material has proven to be very difficult.

An optical memory for a classical finite state machine is considerablely simplier and ultilizes more of the potential of optics. The storage medium in this case need only delay the information for 1 cycle. The other advantage is that no addressing mechanism is needed and thus the potiental parallelism is not restricted.

A Proposed Optical General Purpose Computer

As an example of one of the possible applications of digital optics a simple optical general purpose computer is presented. It is implemented along the lines of a classical finite state machine as shown in Figure 5. The basic approach involves decomposing the structure of a classical finite state machine into a logic unit, an interconnection array, and a latching unit.

An Optical Logic Array

Any digital circit can be constructed completely out of NOR gates. Some NOR gate equivalents are shown in Figure 7. The topology of a digital circuit can be rearranged to form a column of gates as shown in Figure 8. The interconnections of the gates can be organized by conceptually combing the wires to confine the interconnections within a given area as shown in Figure 9. This concept can be extended into two dimensions. The result is a combinatoric logic unit comprising a NOR gate array and an interconnection array as shown in Figure 10.

An optical NOR gate array can be constructed in the following manner. If two or more checkerboard images are projected on a common surface the resulting image will be the OR of these images. If this resulting image is inverted in contrast and thresholded then the intensity of each square of the output checkerboard represents the NOR of the inputs of the square. This is illustrated in Figure 11. Thus an array of NOR gates can be constructed with a non-linear contrast inverting mechanism.

Such a non-linear contrast inverting device can be implemented in several ways. A Hughes Liquid Crystral Light Valve could be used to achieve a cycle time in the hundreths of seconds. The ITEK Optical PROM could probably achieve a cycle time in the order of milliseconds. An array of "optical transistors" implemented with integrated optics can probably achieve cycle times in the order of microseconds.

Interconnection Array

The outputs of these NOR gates are used as the inputs of other NOR gates. This redirection is performed by the interconnection array. The output of each square is redirected to be the input of other squares. This operation can be performed with a hologram. Unfortunately, the hologram for a random interconnect pattern would be quite difficult and quite specialized. A simpler approach is to customize a regular interconnection pattern. Such a regular interconnection pattern is shown in Figure 12. Each square represents an optical NOR gate. The top circle in each square is the input while the bottom circle is the output. The output of a NOR gate with a location (X,Y) is distributed via the latches to the inputs of the NOR gates at locations (X,Y-1), (X+1,Y-1) and (X-2,Y-1). The hologram to accomplish this associates a reference point source (X,Y) with three object point sources located at (X,Y-1), (X+1,Y-1), and (X-2,Y-1). Such an optical distribution system would distribute the outputs of any number of inputs in parallel. This regular interconnection can also be accomplished with beam splitters and prisms, however the holographic

approach would have greater optical efficiency.

To implement a particular interconnection pattern the regular interconnection pattern must be customized. This can be accomplished in two ways. A mask as shown in Figure 13 can be used to block the outputs from particular NOR gates before they can be distributed. Another approach is to prevent particular NOR gates from having any output by forcing their inputs "high". This is shown in Figure 14. This last approach is considerably more flexible since the customizing interconnection pattern can be accomplished by projecting a pattern on the input of the optical NOR gate array.

Some examples of this projection customizing are shown in Figure 15. Merges, forks, cross overs, left, right, and straight are some of the connections which are possible. These are sufficient to form a complete connective set.

Optical Latches

The only requirement of the latches in a classical finite state machine is that they can store the output of the combinatoric logic unit for one cycle. This allows the output to be used as part of the input of the next cycle. In some early prototypes of conventional computers this was accomplished with simple delay lines. The necessary amount of delay is determined by the speed of the logic unit, which in this case is limited by the non-linear contrast inverter. The overall structure is shown in Figure 17.

If the delays required are too large to be accomplished by a simple optical delay then a ring approach as shown in Figure 16 can be used. The output of each combinational logic unit (interconnection and logic arrays) is used as the input of another such unit. N such units, each providing a little delay, are arranged in a ring. If the ring is sufficiently large then each non-linear contrast inverter will have a sufficient time to recycle.

Input and Output

The basic hardware is determined by the customized interconnection pattern. The program or software of the system can be input as a pattern which further customizes these interconnections. The input data can be entered as still another pattern. The output of the system is just a portion of the outputs of the NOR gate array. Thus the hardware, software, data, and output can all be handled in parallel.

Overall Architecture

An example of a section of such a general purpose optical processor is shown in Figures 18 through 22. Figure 18 shows the bias customizing pattern. Figure 19 shows the contents of the latch. Figure 20 shows the OR of the regularly distributed latch contents and the external inputs. Figure 21 shows this ORed with the customizing pattern. Figure 22 shows the result of a NOT operation (contrast inversion and thresholding) on this array. This result is latched and used as the input of the next cycle.

On each new cycle the signals propagate to a new row. The first row accepts a new set of signals on each new cycle. The system is naturally pipelined. If there are N rows then N sets of signals are in partial states of processing at any given time. This pipelining facilitates very large throughputs.

Other Computational Structures

The signals in the simulation shown in Figures 18 - 22 propagate through the system much like "optical bubbles". A 25 by 25 array of optical NOR gates can be used as a shift register if interconnect patterns of (X,Y) -> (X+1,Y) and (X,Y) -> (X-24,Y-1) are used. As another example, interconnection patterns of (X,Y) -> (X+1,Y) and (X,Y) -> (X-24,Y-8) will produce a 75 element 8 bit wide shift register.

The formalism of a classical finite state machine can also support non-Boolean based computing structures. One example is a number theoretic processor[HUA3]. Another example is the customizing pattern shown in Figure 18. This pattern implements an "Amida Kuzi" interconnection network [HUA2] as shown in Figure 23. This network is one of the basic building blocks of a number theoretic processor pipelined by moduli.

New computer architectures which better ultilize the parallelism of these processors should be explored. Current computers are designed around the Von Neumann bottleneck. Duplicating these machines would just perpetuate this bottleneck. Expressed in other words the difficulty is finding problems with a sufficient amount of parallelism to exploit. One possiblility is the processing of images with the two dimensional architecture simiar to the TSE computer [SCH].

Summary of the Proposed Optical Digital Computer

The proposed general purpose optical computer is very simple. It consists of a logic array, an interconnection array, and a delay. The system is designed to exploit the parallelism of optics and non-planar propagation. The fabrication is simplified since the wiring is replaced with a customized regular interconnection pattern. The packing density of optical NOR gates should be greater since each optical NOR gate is simpler and needs no planar interconnections. The communications paths are non-planar, parallel, high bandwidth, and need no termination. The only detection and restimulation occurs within the non-linear

contrast inverting device. The inputs and outputs are parallel and can support very large throughputs. The system is also naturally pipelined.

The practical limits on the array size and cycle times have yet to be determined. The throughput of this proposed general purpose optical computer is the product of the number of optical NOR gates in the array times the cycle time of the non-linear contrast inverting mechanism. A system with 200,000 NOR gates and a cycle time of 1/100 of a second would produce a throughput of 20 million bits per second which is comparable with conventional computers.

Discussion

The diverse nature of the state-of-the-art of digitial optics demonstrates that many optical phenomena can be used to perform digital operations. These devices will remain just curiosities unless they can fulfill some some basic need. Mimicing electronics is not enough. It is crucial for these devices to use some of the unique advantages of optics such as its parallelism or non-interfering propagation. This is not a easy thing to do since any interface with electronics will squelch the parallelism of optics.

A simple optical digital computer is presented. The architecture is unusual in that it is fully parallel. This gives optics and advantage since it would be very difficult to implement such a processor with electronics. The computer consists of a NOR gate array, a latching mechanism, and an interconnect array. Each of these sections is conceptually quite simple and can be implemented by many types of optical devices. The intent of presenting this processor is to start people thinking by demonstrating some of the techniques, problems, and possibilities of using digital optics.

Acknowlegement

This work was supported by the Air Force Office of Scientific Research under grant AFOSR-77-3219.

References

[BAC] John Backus, "Can Programming Be Liberated From the Von Neumann Style? A Functional Style and Its Algebra of Programs", Communications of the Association of Computing Machinery, vol. 21, pp. 613-641, August 1978.

[BAS] N.G. Basov, W.H. Culver, and B. Shah, "Applications of Lasers to Computers", Laser Handbook, edited by F.T. Arecchi and E.O Schulz-Dubois, North-Holland Publ. Co., 1972.

[CAS] D. Casasent, D. Psaltis, and D. Neft, "Digital Multiplication and Addition by Optical Convolution", Proceedings of the 1980 International Optical Computing Conference, vol. 232, SPIE, Bellingham, Wash. April 1980.

[CIN] I. Cinderich, A. Tai, J. Fienup, and C. Aleksoff, "Proceedings of Optical Processing Systems", SPIE vol. 185, Bellingham, Wash., 1979, p. 2.

[COL1] S. Collins, J. Mckay, and C. Vick, Digest of Papers of Compcon Spring 78, IEEE Comp. Soc., San Francisco, IEEE, 1978, p.198.

[COL2] S. Collins and M. Fatehi, "Optical Logic Gates Using a Hughes Liquid Crystal Light Valve", Proceedings of the 1980 International Optical Computing Conference, vol. 232, SPIE, Bellingham, Wash. April 1980.

[GAR] E. Garmire, J. Marburger, and S. Allen, "Incoherent Mirrorless Bistabel Optical Optical Devices", Applied Physics Letters, vol. 32, p. 320, March 1978.

[GUE] C. Guest and T Gaylord, "Parallel Truth-Table Lookup Digital Holographic Processing", Proceedings of the 1980 International Optical Computing Conference, vol. 232, SPIE, Bellingham, Wash. April 1980.

[HUA1] A. Huang, "Implementation of a Residue Arithmetic Unit Via Optical and Other Physical Phenomena", Proceedings of the International Optical Computing Conference, April 1975, Washington D. C., IEEE Cat. no. 75 CH0941-5 C.

[HUA2] A. Huang, Y.Tsunoda, J.W. Goodman, and S. Ishihara, "Optical Computation Using Residue Arithmetic", Applied Optics, vol. 18, no. 2, pp. 149-162, January 1979.

[HUA3] A. Huang and J. W. Goodman, "Number Theoretic Processors, Optical and Electronic", Optical Processing Systems, SPIE Vol. 185, May 1979.

[JOH] E. Johnson, L. Risebrough, A. Lampicki, and H. Samelson, Applied Physics Letters, vo. 26, p. 444, 1975.

[LAN] R. Landuaer, Optical Information Processing, vol. 1, (Proc. of US-USSR Seminar, Washington, D.C., 1975, ed by Nesterikhin, Stroke, and Kock), Plenum Press, N.Y., 1976, p.219.

[LEE] S. H. Lee, Optical Information Processing, vol. 2, (Proc. US-USSR Seminar, Novosibirsk, 1976, ed. Barrakette, Stroke, Nesterikhin, and Kock), Plenum Press, N.Y., 1978, p. 171.

[MNA] E. Mnatsakanyan, V. Morozov, and Y. Popov, "Digital Data Processing in Optoelectronic Devices (Review)", Sov. J. Quantum Electron., 9(6), June 1979, (transated by American Institute of Physics).

[NIS] P. Nisenson and S. Iwasa, "Real Time Optical Processing with $Bi_{12}SiO_{20}$ PROM", Applied Optics, Vol. 11, No. 12, Dec 1972, p. 2760.

[ORL] L. A. Orlov and Y. M. Popov, Sov. J. Quantum Eectron., vol. 4, 12, 1974, (translated by the American Institute of Physics).

[SCH] D. Schaefer and J. Strong, "TSE Computers", Proceedings of the IEEE, Vol. 65, No. 1, January 1977, p. 129.

[STO] F. Horrigan and W. Stoner, "Residue-based Optical Processors", "Proceedings of Optical Processing Systems", SPIE vol. 185, Bellingham, Wash., 1979, p. 19.

[STR] T. Strand, A. Tanguay, A. Sawchuck, P. Chavel, D. Boswell, A. Lackner, and B. Soffer, "Optical Computing with Variable Grating Mode Liquid Crystal Light Valve", Proceedings of the 1980 International Optical Computing Conference, vol. 232, SPIE, Bellingham, Wash. April 1980.

[TAY] H. Taylor, Applied Optics, vol. 17, p. 1493, 1978.

[TOR] H.C. Torng, Introduction to the Logical Design of Switching Systems, pp. 180-196, Addison-Wesley, 1966.

CLOCK SKEW

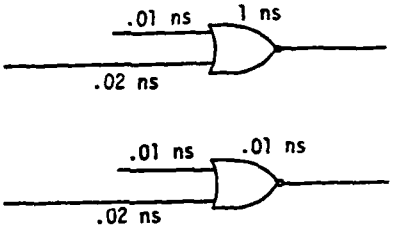


Figure 1 Clock Skew Problem

BANDWIDTH

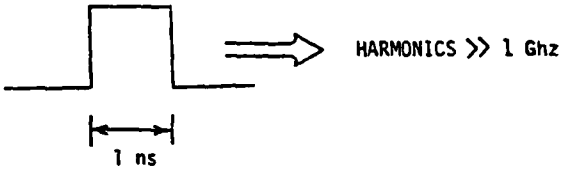


Figure 2 Bandwidth Problem

INTERCONNECTIONS

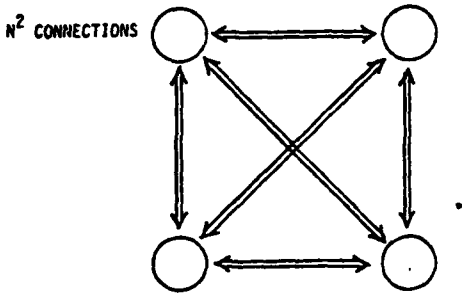


Figure 3 Interconnection Problem

VON NEUMANN BOTTLENECK

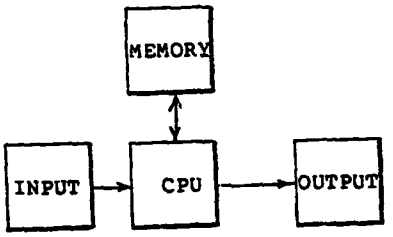


Figure 4 A Von Neumann Computer

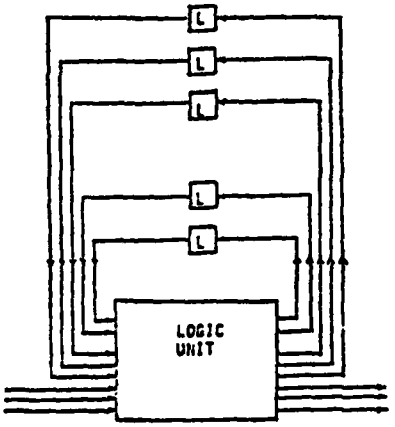


Figure 5 A Classical Finite State Machine

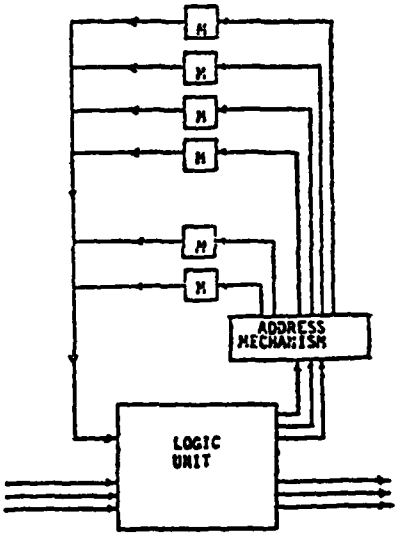


Figure 6 A Modified Finite State Machine

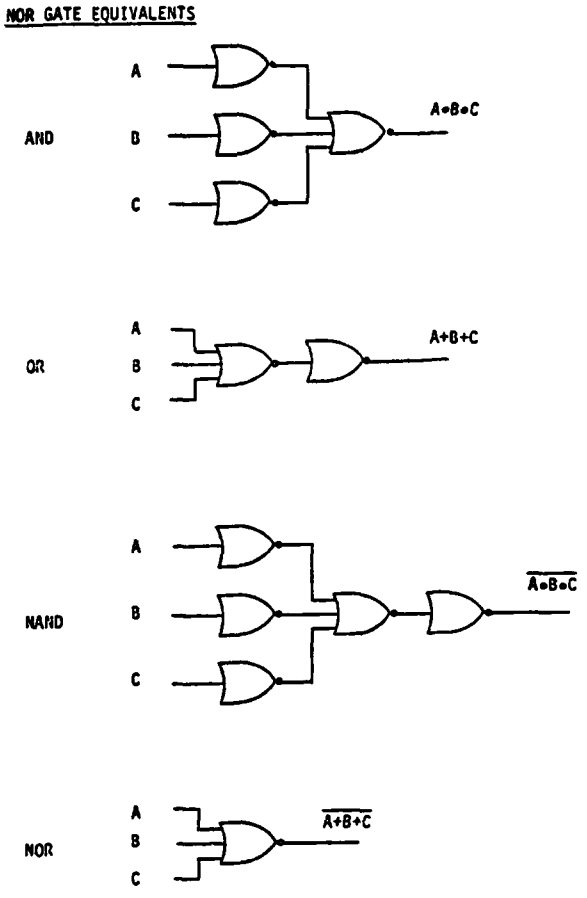


Figure 7 NOR Gate Equivalents

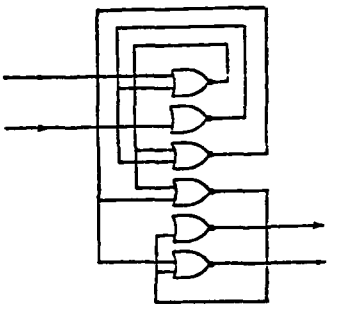


Figure 8 Column of NOR Gates

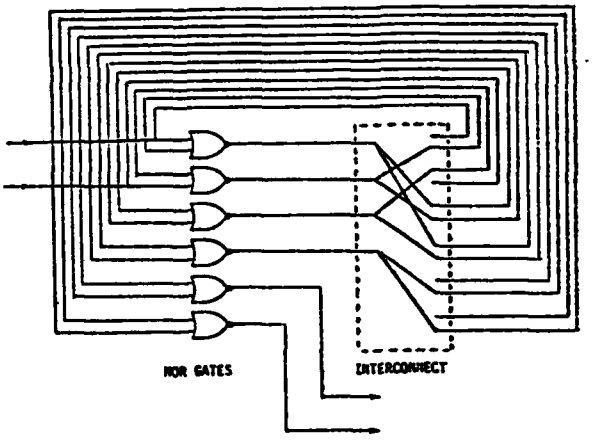


Figure 9 Combed Interconnections

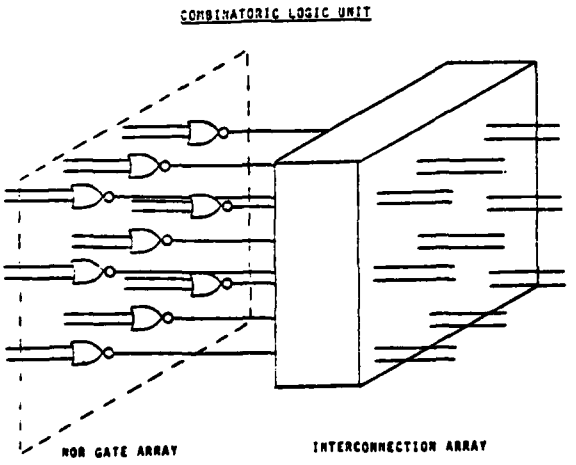


Figure 10 Combinatoric Logic Array

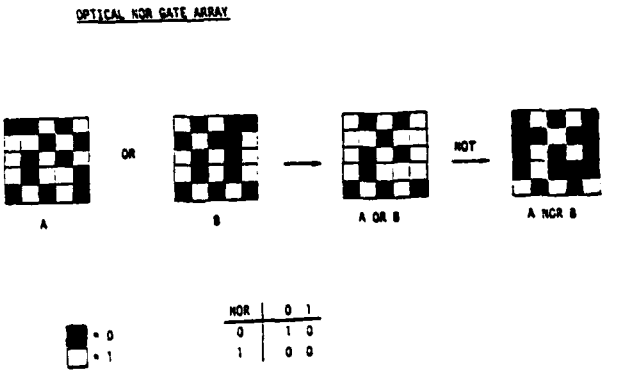


Figure 11 Optical NOR Gate Array

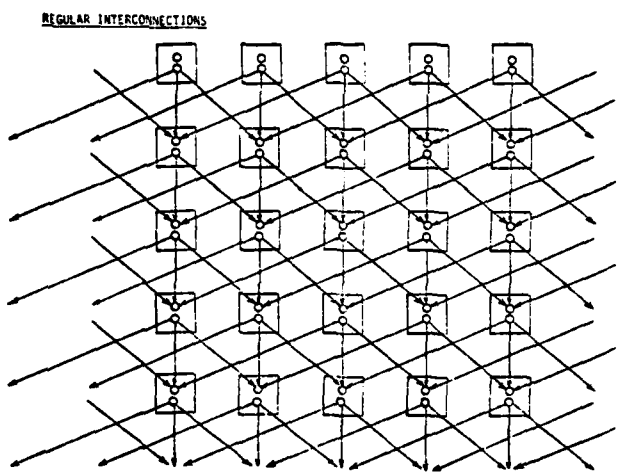


Figure 12 Regular Interconnection Pattern

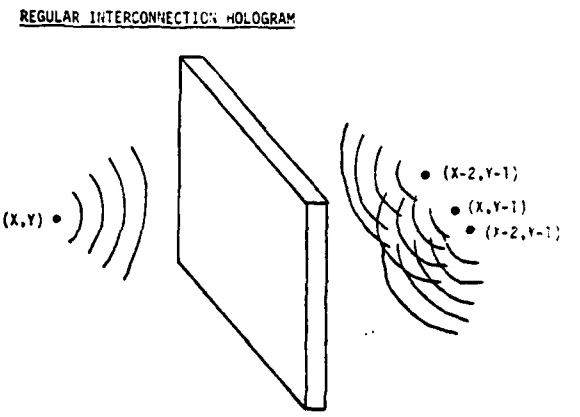


Figure 12A Regular Interconnection Hologram

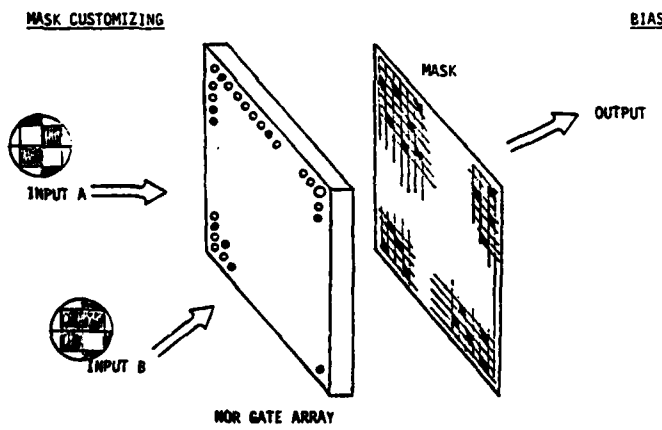


Figure 13 Mask Customizing

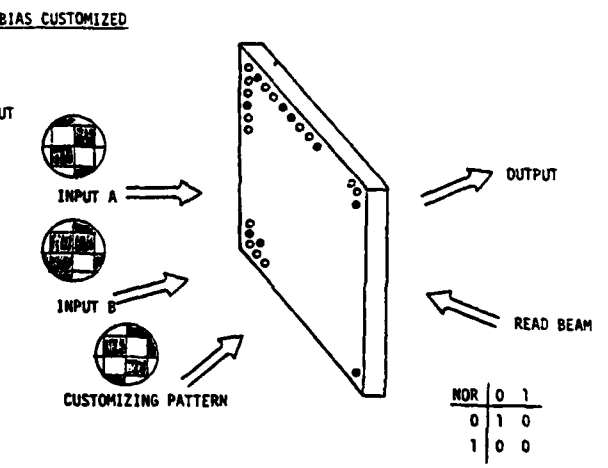


Figure 14 Bias Customizing

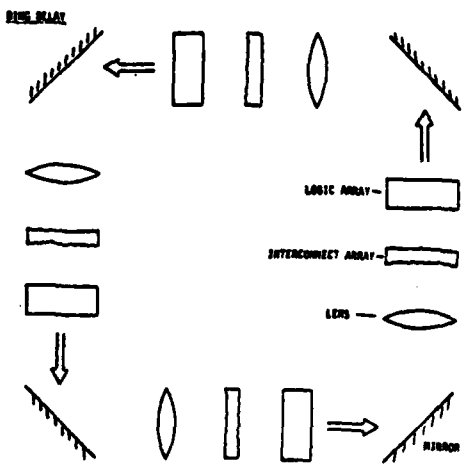


Figure 16 Ring Delay for Latching

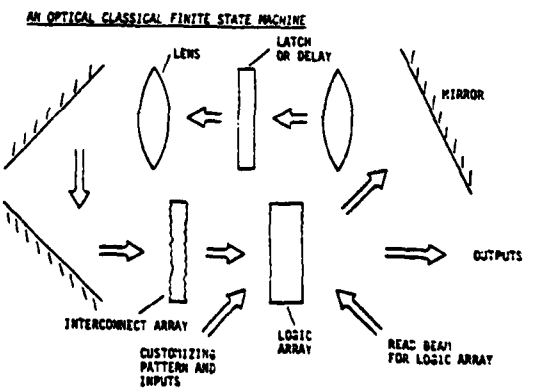


Figure 17 An Optical Finite State Machine

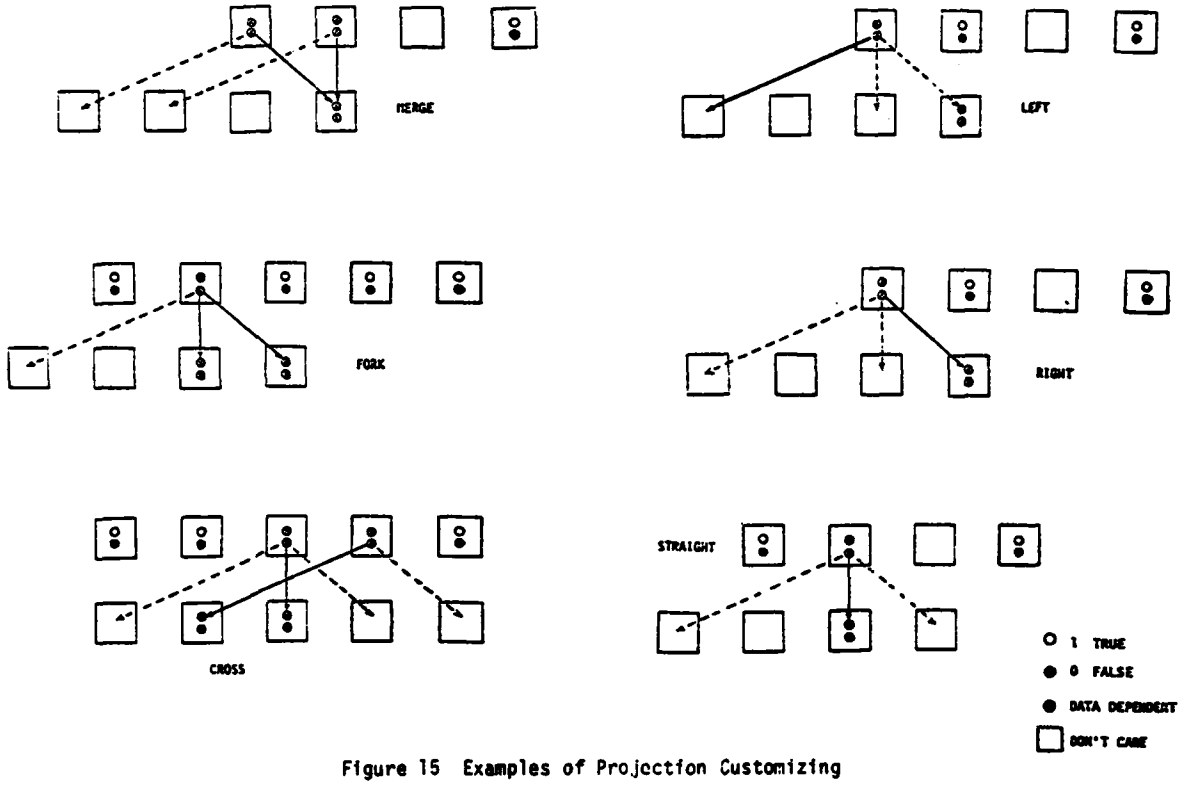


Figure 15 Examples of Projection Customizing

BIAS CUSTOMIZING PATTERN

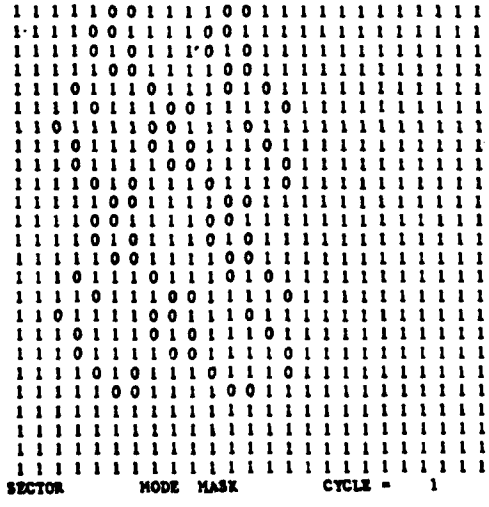


Figure 18 Bias Customizing Pattern

CONTENTS OF THE LATCH

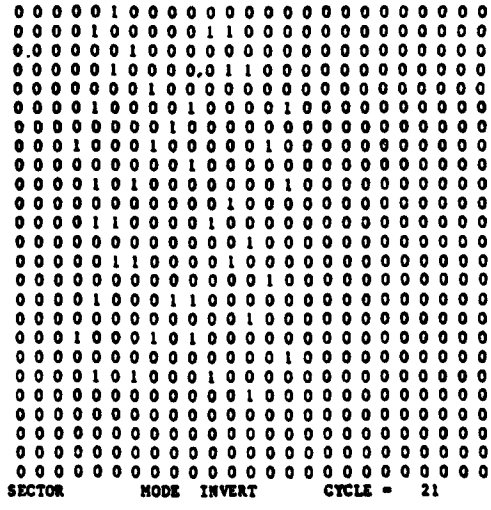


Figure 19 Contents of the Latch

THE OR OF THE REGULARLY DISTRIBUTED LATCH CONTENTS AND THE EXTERNAL INPUTS

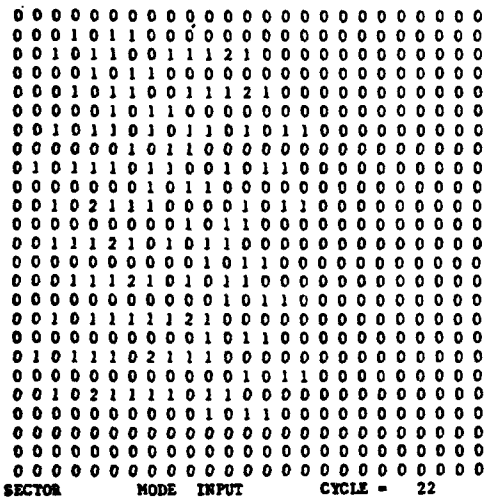


Figure 20 The OR of the Regularly Distributed Latch Contents and the Inputs

THE OR OF THE CUSTOMIZING PATTERN

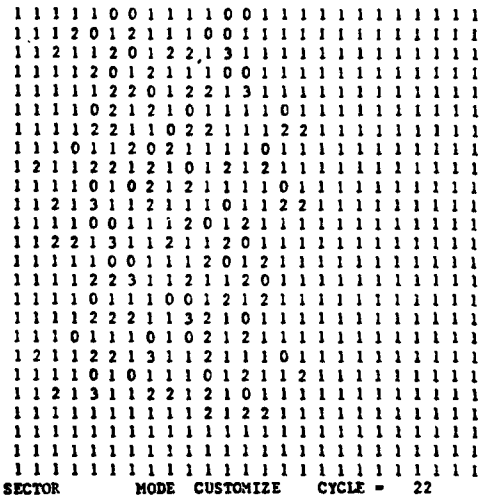


Figure 21 The OR of the Customizing Pattern and Figure 20

THE INVERTED OUTPUTS OF THE NOR GATES

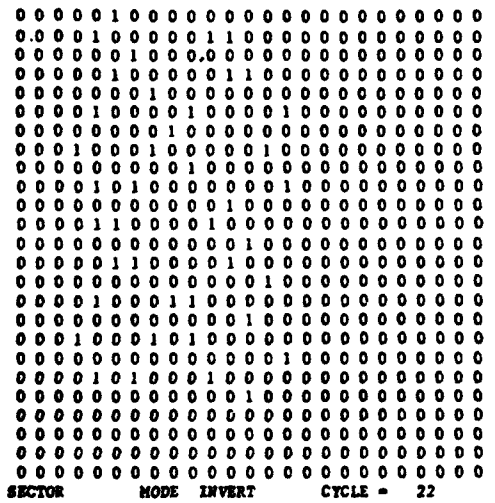
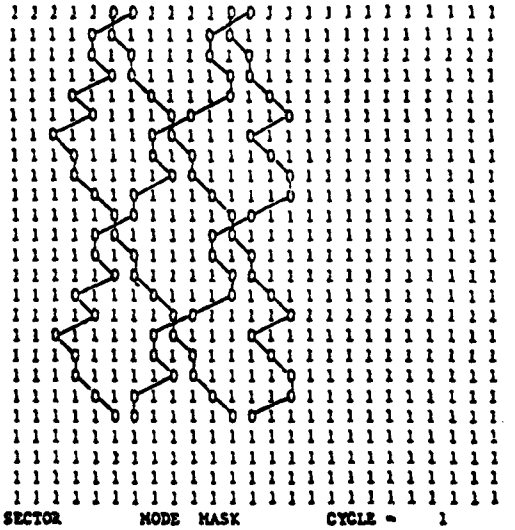


Figure 22 The NOT of the Array

AMIDIA KUZI CUSTOMIZING PATTERN



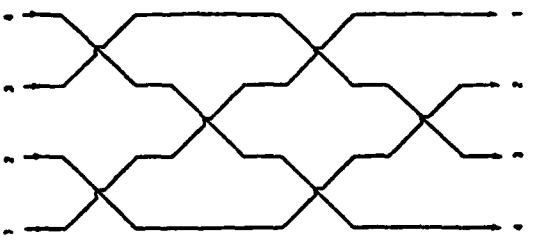


Figure 23 "Amida Kuzi" Customizing Pattern and Interconnection Network

Discussion (Alan Huang; Discussion Leader: Stewart A. Collins)

C. You referred to the customizing masks with which you essentially set the state of the system. I think of this as an optical micro-programmer.

R. Yes, it is an optical way of specifying the hardware, and on another level it is a micro-programmer.

Q. Have you considered problems that people back off from doing with conventional technology? For example, in designing X-ray tubes, the computer solution of Poisson's equation in 3-D is prohibitive in time and money. The same problems come up whenever you're trying to solve a partial differential equation.

A. In solving a diffusion equation or heat equation, the procedure is something like balancing the mean between the extremes. I think that is a very nice local operation if you can work out the programming of interconnects so each cell will do that type of thing.

R. It might be a simple program compared to the more sophisticated algorithms that are used now, like the finite difference technique which recasts the whole problem as a matrix problem. It may be possible to use a simple relaxation technique and just let the optical computer cycle through until a solution is obtained. It wouldn't even be necessary for everyone to have one of these computers if it was possible to use a time-sharing system and have a few of these at centers that could afford them.

Q. You've been working on this general purpose optical idea for a year or two now. Have you developed some thoughts as to what its particular strengths might be?

A. At first I had to use a pipeline architecture because I was watching very few bits change on a simulation of a regular finite state machine. What I was doing was re-creating the bottleneck in software which I had spent all my time avoiding by using the parallelism of optics. Also, it was a little too tight in terms of the competition, so I stacked the deck and found things that are very difficult for them to do. One such thing is the case where you want to do a weighted mean of values that are 2 or 3 neighbors distant in an array. This would require a prohibitive number of interconnects for a parallel electronic processor. I'm staying more toward the interconnect problems because that is where optics has an advantage since I am doing these interconnects in space.

Q. You've designed your operations around the NOR gate. If you had your choice of 3 or 4 different logic operations, which would you prefer?

A. I did the NOR gate because optically you get the OR for free, so the basic operation you have to get is the NOT operation, which is very difficult. I would also be satisfied with the NAND operation. For all other operations, you need a combination of two. A computer designer is perfectly happy with NOR's.

Q. If you had a way to do the interconnections, would you be able to do arbitrary algorithms?

A. Yes.

Q. Would you have to change the interconnection pattern?

A. No, the interconnect pattern is fixed, that's simple and primitive. You would just project another customizing pattern.

C. Keyes at IBM has done a famous analysis of different types of logic elements from the point of view of their limitations. His conclusion was that optical logic was not viable because of the high energy involved and the cost of a photon. He thought that the heat problem which is preventing Si technology from going much farther was even worse for optics. You've introduced a new twist here. You've said that you don't require nanosecond switching times, which might alleviate that problem and get around Keyes.

R. Some people mentioned that they now have integrated optic components that can latch with less than 100 photons of energy, and that is quite sensitive. You can make the liquid crystal light valve more nonlinear and I think you can get some speed that way. One tendency I have is to take the output and feed it back to put some gain in the system, but that is still trading time for energy. The whole architecture is not basing my throughput on switching speed, it's the parallelism.

C. Keyes was talking about an exact mimic of digital logic with single channels and single gates, and parallelism introduces a new dimension.

R. Especially if you limit the number of times you have to re-create a photon. A photon is an electron volt, and that's a lot of energy.

C. In electronics, people started out with analog computers and eventually found out they could get more use out of the logic and dynamic range of digital devices. It seems now we're in the analog stage even in image processing. Could it be that eventually we'll want to shift over to some form of digital parallel optical image processors?

Panel Discussion

Moderator: John Walkup (JW)

Panelists: Bud Vanderlugt (BV), Joe Goodman (JG), Sandy Sawchuk (SS), Harper Whitehouse (HW)

(Key: C = Comment, Q = Question, R = Response.)

JW: Just for the record, the panel consists of Bud Vanderlugt, Joe Goodman, Sandy Sawchuk and Harper Whitehouse. I very much want to encourage questions and comments from the rest of the participants so that the panel not do all the talking. I thknk they would concur with that. It is hard to know exactly where to start. There was a question submitted and this is the only question that was put in the box. Maybe it is a good place to start:

Q: What is the advisability of pursuing optical binary logic techniques in the presence of the electronic digital capability?

BV: My thoughts right now go back to experiences of block oriented holographic memories. The question I have is, even if we can get the switching speeds and get enough of the devices developed, is the volume that is dependent on free-space propagation that may be required in some of these applications going to be such that we can package it in a viable size and with the kinds of power consumptions, etc. that are competitive with the digital approaches. I think that has a lot to do with the way in which the system architecture is structured. I know that when we look at holographic memories with block oriented random access structures, we can get very nice packing density in the Fourier plane, but when we consider the tremendous volume that is required for the free-space propagation to get the beams there, do the Fourier transformations, etc., the total system size will be large, i.e. the volume packing density is not all that high. I wonder whether the same issue may be applicable here.

JG: It seems to me there are two directions one can go here. One is towards making extremely fast logic elements, and indeed there are people working on that. They have hopes of far surpassing what can be done with GaAs technology. I think Alan Huang put his finger on the other kind of approach, which is not to fight the speed battle, but to try to employ the massive parallelism that we have available, in trying to achieve a large number of operations per second through parallelism rather than through basic, fundamentally superspeed devices. I think that is a good insight.

HW: I have feet in both camps. I'm working with the Navy on its VHSIC program which is going to be the panacea of five years from now for all of digital processing. We can't even solve all current Navy systems from the projected components. Part of the problem is that much of the VLSI chip is typically taken up with internal interconnects, thus only the most regular arrays can be implemented in VLSI. The pin limitations on the chips are such that much beyond 128 pins/chip, people will begin to not see how they are going to package it and how they are going to test it. Also with the very fine line geometries, people are beginning to become concerned about how these are going to behave with soft errors and cosmic ray-induced errors. I feel that the solution to the problem is not to try to make things work faster. People are working with Josephson junction devices to get speed. When you ask as we did on the program, how would Josephson junction devices solve the same Navy problem as is being addressed by the three optical computers, we find that even with the existence of the Josephson junction computer, postulating that the computer now exists, we then had to postulate a Josephson junction array processor to go with the Josephson junction computer in order to be able to solve the problem. So speed (per se) does not do it. I believe it is the exploitation of the 2-D structure and addressing those algorithms which are of N^2 today, just as the FFT addressed the N^2 operations problem of the DFT. But before one starts on that we have a precursor that has come out of Stanford University for the inversion of a class of matrices which now takes N^3 operations. They have now been able to show algorithms with $\alpha^2 N(\log_2 N)^2$ operations, however, α unfortunately ranges between N in the worst case, to 1 in the best case. So we have to fight. One is not sure exactly; but for many problems α is a small number, and notice that we have reduced N^3 operations to $N(\log_2 N)^2$. So even though optical processing can bring N^2 processors to bear, the algorithm people are not sitting back idly, and Stanford University is addressing the problem of VLSI through algorithm research. So I think it is going to be nip and tuck.

SS: I don't know if there is a great deal that I can add to this except to say that there is a lot of work that remains to be done in studying the architecture of both parallel digital processors and the architecture of parallel optical processors. And people like Alan Huang and others have just begun to think about what you can do with the parallel optical logic of residue arithmetic arrays. As Alan said, and I agree with this, I think that one should not be impressed with speed at the moment or be so concerned with it. But instead think what can be done with parallel processors. Try to identify those specific problem areas that would benefit from a parallel processor. To me it seems that one of the areas is in dealing with data that is page organized to begin with. Such data is, for example, images of any kind, 2-D arrays as sonar, radar, whatever you have. So these are natural candidates for some sort of parallel processor and we should try to examine these problem areas, find out what problems digital people have, and they are considerable. Another thing is, as Harper said, VLSI and other technologies that are being explored now are sort of billed as the universal solid state answer to all these problems, and they really are not, because they have many fundamental problems. Among them is the interconnection problem of the different units. In an optical computer you might

imagine that interconnection is easier. But it does take space, as Bud pointed out. It takes a reasonable amount of space for light to propagate over a Fresnel distance, or whatever it takes to get the Fourier transform. So we might look at perhaps trying to shrink the sizes of these things by combining them with fiber optics and integrated optics. And finally, and this is a wild idea by a guy who does not know much about integrated optics, except we have all heard a lot about 2-D integrated optics, I am wondering what the possibilities are of 3-D integrated optics for doing some of these massive shrinking and size reduction types of things we have been hearing about. This is something perhaps we can come back to later and talk to Tsai about.

JW: Chen (Tsai), do you want to respond to 3-D optics?

R: At this time we have enough problems with 2-D. But I know that in 3-D, one possibility is to have a stack, for example for direction finding. At this moment it seems to be difficult, but in the meantime research progress in materials is being made, IBM is growing thin layers. But it is obviously very far away, and I think at this moment it is very difficult. [Chen Tsai]

SS: I didn't really expect that there would be a very good answer on this. But two other things that I though of; as I pointed out before I tried to show that I really feel that getting thresholds, or a parallel array of thresholds, is a real key to doing digital processing. By digital I mean by optical techniques as opposed to analog optics. I also think that the digital technique is a very big key to accuracy. This is the whole key to an electronic digital computer. The reason why they get the accuracy is because of its thresholding nature. And perhaps we don't want to mimic this exactly with optics, but at least we should be able to do a threshold, and this is a key to doing any kind of logic-decision making, branching, testing, and all the other things that one can do with an electronic computer. Finally the other thing that I want to say concerns iteration. In order to do some of these techniques, like solving a set of equations such as Harper mentioned, you need iteration. This implies feedback and implies memory. You have to be able to store computer results somewhere, bring them back and work on them again and again and again. This is something that electronic digital computers can do and I guess people are thinking about this with optics. But this is again a part of the big architecture question.

C: I would like to endorse the viewpoints that you have mentioned. One is that you should utilize, what was actually expressed by Joe and others, the progress that has been made, and also is in the making, in fiber optic systems. Optical signal processing systems should utilize this progress, that is very important. The second thing to study is what the suitable architectures are to go with these planar structures such as 2-D time integrated correlators. [Chen Tsai]

JW: When we talk about thresholding, and it was mentioned in one of the talks, if we have feedback then we can get some of these functions. There has been relatively little talk about issues such as the optical operational amplifier. I know that Sing Lee has been working on something like that and Marion Hagler here at Tech did a fair amount of analytical work on what the requirements will be. How useful in the opinion of the people here, or how critical will it be, to develop an optical op amp that can work on 2-D data? A lot of things we talked about almost imply that we need iteration, we may need feedback and we need high gain.

HW: I think I would vote for the optical comparator first. By comparator I mean not just a threshold but a threshold for which I can control where the threshold is. I mean it to act like a two dimensional array of comparators.

JW: We know that Adolf has done a fair amount of work on TV optical-electronic op amps and has got very interesting results.

C: You've already got your optical op amp-the image intensifier tube. But the key item you want is an optical op amp which has negative dynamic gain characteristics. [Stuart Collins]

JW: Let me get into a couple of questions that were on this sheet that I filled out. I had two that seemed to go together; What subareas of optical information processing appear most promising for dramatic developments over the next decade: In other words are there 3, 4, or 5 that you would say are right on the verge of breakthroughs or could be on the verge of significant breakthroughs?

JG: I wouldn't like to answer that question, but I would like to rephrase it, because I think in order to answer that question we need a crystal ball and I don't have it. I think a better question would be what new developments do we see that might potentially have something to offer to optical data processing, but at this particular stage for such developments it is likely to be too early to tell. The only way to find out is if you pursue them with vigor for awhile to see what comes out of them and I think I would put into that category right now the various ideas that have been developed for using four-wave mixing for correlations, convolutions, and various things like this. This is basically a dynamic light valve of the sort that performs nonlinear operations. It seems, at face value anyway, to have enormous possibilities, at least to the degree that deserves further investigation. Maybe after looking at it for awhile we will begin to see the defects as we always do after we examine something for awhile, but at this point of time is seems to be an area that really should be looked at carefully.

JW: Anybody else on the panel want to respond?

HW: I think it is precision. I think until the problem of optical precision has been addressed there will never be any rest to the problem of competition from the digital community. Only when that problem is addressed, when we can say, "I will compare my 8 bit digital system with your 8 bit optical system," then we can get down to the size, weight, power, and cost comparison.

JW: Is it precision or is it a sort of throughput times precision or some throughput accuracy?

HW: I think it is precision, at sufficient throughput to solve the problem; excess throughput beyond that required to solve the problem I don't think is needed.

JW: I think that is a task-oriented thing.

JG: I agree with Harper that precision is a key thing that we should be focusing on and certainly we devoted some of our efforts to it at Stanford in the past few years. But I wanted to raise a paradox, i.e., if we believe Shannon, the information rate that is coming out of a processing system is linearly proportional to the throughput, but only logarithmically proportional to the number of resolvable levels. Therefore doubling the number of bits that we have available, there should be a very substantial increase in dynamic range, which in Shannon's terms is only equivalent to doubling the throughput rate, and I can't quite understand why Shannon tells me one thing and my intuition tells me another. I wonder if you have any comments on that.

HW: The intuition loses: I think that this is a problem that should be addressed from the information theoretic point of view. We need to put optical processing on the same basis that I think has been done with super-conductivity. In Conway and Meades' new book, chapter 9, there is an interesting comparison between MOS technology and Josephson junction technology. Based upon quantum physics arguments, which show that as we scale devices, then semiconductor technology, the MOS technology in particular, improves in performance; one type of Josephson junction technology is neutral, i.e. it makes no difference, and another type of super conducting technology loses, actually it is worse as the resolution increases. I think we need to address questions like these, the information theoretic one, and we need to get down and address them at university level physics. I would like to see optical processing plotted on the same curve that has semiconductors and Josephson junction devices. I think we have to have all three put together from a physics point of view, then we can answer some of these comparison questions.

C: You talked about addressing accuracy in optical processing. That has, in the last year, really evolved as a major criteria for digital processing. People are trying to build them smaller and faster. One thing they have found, that up until last year they have overlooked, is the quantization error that occurs in that processor which is dependent on the structure of that processor. So now when you are comparing digital processors you must address this quantization error problem. So even in digital processing with IC's, accuracy is still a major consideration.[Bill Sander]

HW: At this point I would like to make a comment on that. Something recently has come out, i.e. people addressing this problem have made a comparison between modern spectral processing techniques with the old fashioned ones. One can show theoretically that under many circumstances this should not make any difference. Yet significant improvements in Navy systems are possible. When you apply the techniques that should not work mathematically any better than the other ones, they do. Thus recently there has been, out of ONR, a new set of opportunities in addition to contract research opportunities which have been available for many years. Starting one year ago there are SRO's, select research opportunities. This is the second year. And for those in the optical signal processing community two of them are of interest. One addresses non-Gaussian signal processing, the other addresses non-stationary signal processing. I would like to mention to those of you who are looking for things to submit, take a look at these select research opportunities. You may find something for optics.

JW: Does anybody in the audience want to respond?

C: I just want to make a comment. Over the last 3 years there has been diversity in processing techniques, in architectures. From what we heard yesterday, when you go to talk to people that are in the digital world, they are kind of locked into certain architectures of computers. I think that maybe this community should be trying to keep the diversity of architecture of optical computers coming on. Lots of times you look at them and they end up being essentially some kind of Fourier transforms and correlations and you are pretty much stuck with them. Try to put them into different concepts and to keep novel new architectures alive. Microprocessors, I think, have made many people come and look at strange architectures for computing systems and it is strange in the sense that they had not thought about this when they were thinking about large scale, big throughput systems and now I think the optical community should try to keep looking into these kinds of areas. [Terry Stalker]

JW: Let me ask a question related to who tackles what problems in basic research, so to speak. It seems like on the one hand there is a need, especially I think it is perceived by the DOD community, to have the research that is done address, or at least have implications for, the needs that are sensed by the DOD. On the other hand I hear people like Bob Guenther saying you've got to look at the basic physics, what are the fundamental limitations? When you start talking about fundamental limitations it seems like you really have to go back to very basic approaches which may appear to be a kind of work that is rather uncoupled with any needs, or loosely coupled with direct mission oriented kinds of things. I think this is a dilemma sometimes for the people in the universities who really are the people who should be, or maybe are, interested in pursuing some

of these basic problems. Very basic types of things that make good theses and dissertations, whereas the things that are a little farther along the development chain are not always as suitable for theses and dissertations. Is there a dilemma there? I see the Army people shaking their heads saying no.

C: There is no dilemma there. We are the ones who are responsible for identifying the roles of the very basic research. And that is why we say contact us-phone call, letter, proposal, whatever. If we don't see the need for that work ultimately, or even in ten years from now, we will tell you. And you have an opportunity to argue your point of view. But it is really our job to do that relevance evaluation, not yours. We want to identify the areas that we can tell you that we are interested in; that doesn't mean that it should exclude anything, that's not what we talk about because we like to hear new ideas also. Maybe something we have not thought about. [Bill Sander]

C: We aren't flooded with a lot of novel, basic research proposals. [Bob Guenther]

C: There are some things in basic research that the Army does not fund. We don't fund astronomy, we don't fund elementary particles, we fund very little work in plasma physics. But otherwise we are even given the latitude that, if we can't even think up an application, we have 25%, roughly, that we can say 'this is the one that we think is worth a risk.' It is just good science, we do have that option. But we just don't get flooded with a lot of ones that we have to come to that kind of decision. [Bob Guenther]

JW: There is an interesting question as to why you don't get flooded.

C: We are, all the OSR's, are a service organization to the different branches of the DOD, and NSF is the government's response to basic research and support. So we shouldn't see a lot, maybe, but there are parts of basic research that all of us that work in the OSR's would give our right arm to see in our program. [Bob Guenther]

HW: May I point out that there is one player missing from this discussion, and that is ARPA. Somebody commented that in comparing 6.1, 6.2, 6.3, there is an order of magnitude of increase in available dollars as one goes towards advanced development. While we all like to think that we are doing fundamental research, we often have to take 6.2, or exploratory development dollars, in order to have enough money, and then divert a fraction of those development dollars into supporting research at the same time we are doing development. Don't overlook ARPA. ARPA has the dollars, the interest at the 6.2 level of development, but they will also work with you, and let you take and redirect a portion of their money into research.

C; They have the 6.1 money?

HW: They do also. But they also let you take 6.2 money and direct a portion of that, so that you can get a single contract from ARPA, which tends to be a bit more applied, and then you can turn to a university and support some student with a small fraction of that money. They are not represented at this meeting. So at least somebody ought to think about them.

C: Basically this is true in any of the DOD. You can spend any of the higher level money for basic research; you can't spend basic research money for higher levels. [Bill Sander]

JW: I wondered if that would be a controversial question. I had hoped that we could get Al Schutzman and Norm Kaplan here. In fact they were invited and I think they were giving some testimony before Congress, so they were unable to come. But certainly NSF is an important place to look for basic research funds.

JG: I had just a comment. I am convinced that it is a delightful situation that places like ARO are really looking for basic research proposals. But from the university point of view, and over my professional lifetime, there seems to be a wave, it goes up and it goes down, and I can only too vividly remember the time when the Joint Services Program for example did not want to be bothered with anything that did not have a very clear relevance to some particular mission. And I think a lot of us who went through that period still have that memory in our minds and may hold back from proposing anything that is really too divorced from either building a piece of equipment or some kind of direct relevance.

JW: I will just use this as an illustration; but it seems like, when you are talking about areas like statistical optics, there are relatively few people working in applying these areas who are really looking at fundamental limitations and the tools are there to answer some of these questions.

C: There are two research programs in our programs that are directed towards those kinds of questions. One about coherence theory and there are one or two directed towards statistical problems in optics. You just don't see them because in a meeting like this you don't bring up someone's theory of propagation. [Bob Guenther]

C: I might make a plug for the sensors and devices as areas needing additional investigation. I think the one key important issue we have not addressed is the materials research. After all, device research boils down to materials, and the interaction of light with matter, and there are various effects, like electro-optic effects, the acousto-optic effect; the magneto-optic effect is hardly used except in magnetic bubble domains and that kind of thing. There are probably other effects out there that might be useful for optical information processing that we have not addressed. And I would like to see more emphasis on fundamental

physics of interaction of light with materials. Can we get a larger index change for a smaller field, should we dope the materials and what we should dope them with? I think it is very limited. Lithium niobate and lithium tantalate and KDP seem to be the only materials that we ever use in all the discussions I've heard in the last two days. There are biological materials out there, there are large molecule proteins that I know seem to have some optical properties that might be useful. But none of us are looking at those things. A proposal in such an area might look farfetched and might look more biological than the Army might want it to look, but I think that we have to think in a much wider sphere of materials, both in sensors and detectors and both in devices for modulation of light and processors. [Cardinal Warde]

C: I would like to return to what Harper mentioned about precision. One thing that struck me about reading optical journals about image processing is that we always see pictures that are being enhanced or level sliced or equidensitometry, including some of our own work. But the heart of any work is mentioning what are the input level plus output levels put in quantized form. It seems that many papers are not talking about precision at all. Most people don't even document bits, 12 bits, 15 bits, it's difficult to even narrow down to 2-3 bits. I think we should all be more careful about all the pictures, what we mean, what kind of input-output we have. One thing I try to use is the transmittance because that is an input-output which is going to be independent of input intensity, i.e. you have a ratio. [H.K. Liu]

HW: We have heard that not a great deal of precision is needed in optics and I think that is often brought about by considering visual images. As soon as you go from visual scenes to infrared detection, much larger dynamic ranges are required, then when one goes to radar imaging one can get very high dynamic ranges. And so one man's problem for which 6 bits is more than adequate may be unacceptable to another man's problem. Just as we have microprocessors at 8 bits to the Cray I computer at 64. I think that we need optical systems which are capable of handling different dynamic ranges for different problems.

SS: In the past, historically most optical systems are built for visual uses. In the past the human is the guy at the output and he is the final sensor. And with human beings you can get away with 5 or 6 bits. But now if we are asking processors to do anything nonlinear, and this includes some of the interesting things like homomorphic filtering for one example, you've really got to have much more accuracy, or else we are going to wind up with results which are just a lot of noise.

SS: One other thing that I want to talk about has to do with what Bob said, namely, why people don't get more fundamental proposals for looking at these things. First of all they are very tough problems and they require probably development of a totally new theory, and work in statistical optics is not developed enough to answer some of the fundamental questions. What would be nice would be to make some sort of an analysis of the limitations, the physical limitations; I sort of look at limitations two ways. One is the physical limitation and the other is the practical limitations. The physical limitations are getting down to the quantum level. How well can you measure light? How well can you measure the signals that are passing through systems? In order to do this we need a lot of statistical tools, some of which have not been developed. And then only after you have these available can you answer more practical questions like what are the noise limitations, the bandwidth and the power, size, and the information throughput, things like this. So there are lots of questions that can't be answered because the tools are not available.

JW: I want to talk about devices and the development of devices. Armand said earlier that if we thought that optics was not going anywhere fast enough, or as fast as it should, we can blame it on the spatial light modulator people. But it turns out that there are only about, in the universities, three or four that we can blame it on, so, I think there is a problem here, there are a couple of problems. There are not enough people working, as Cardinal (Warde) said, for example in the universities, on development of the basic devices. The other issue is how do you get the spatial light modulators, which are very expensive, in to the hands of the people who are doing the advanced techniques development to broaden the scope of the things that optical computers can do, when a lot of us cannot afford them.

SS: Let me answer that. One thing is that I really want to stand up for people who do basic research on the properties of interaction of light with matter. There are a lot of questions out there that need to be answered before we can decide to put a lot of money and a lot of resources on new kinds of light valves. There may be one hundred different ideas you can have on how to build a light modulator. But you don't have unlimited resources, you better pick two or three good candidates by intuition or gut feeling or something like that and then put your resources there. And that is an extremely risky business because the development time is very, very long, it is very slow and has, just by the fact that you have enough money for three projects and there are one hundred candidates, a low a priori probability of success, and the results are very, very slow in coming. The experimental device research is very expensive and very risky. So again you have to use your intuition to modify these a priori probabilities and hope that you come up with something good. But I really want to stand up for doing basic device research, and especially as Cardinal (Warde) said, a lot of questions on the interaction of light and matter are just unknown and unanswered.

C: I would like to put in a plug also for materials research. I feel kind of naked in the stuff I do, like building castles in the sky, and what makes them eventually touch ground is the materials thing. And I feel kind of frustrated because they are doing research in a certain area and they have no idea of what other areas to go into. You can't really start another area because you don't have the devices that will support it. But I think this cycle has to be broken in either of two ways. One is that the systems people really have to stick themselves out and fantasize what is needed and then the device people can move in that direction. Or also the device people can say, "Look, here are these unique phenomena," and find some way for using

it. But I think a lot of this cycle is determined by inertia and, like the Hughes liquid crystal light valve, it is being driven by TV projection criteria and everyone is tagging along. That is where the money is and that is what drives it, but it forces the rest of us in the same direction. We tried to look at the problem independently but economics always drives us to what is available. [Alan Huang]

C: In conjunction with support for research in materials and devices, I endorse that it is needed. But I'd like to take this opportunity to address a question that was raised yesterday by Joe Goodman and others, i.e. how can we induce a larger change in the index of refraction. I indicated yesterday, some progress has been made in lithium niobate substrates and lithium niobate is important because of the fact it has good electro-optical and acousto-optical properties. I indicated that an employee of Bell Labs has demonstrated a change of index of refraction of 0.13. And remember that the index of refraction of lithium niobate is 2.2 and that is six times more than the conventional in-diffusion and I think that is important. Another characteristic, of this is that this is by ion exchange, namely exchanging an ion of lithium with silver or thorium. This is a very simple process. What I want to add to this information is that they have observed that with this kind of ion exchange method, the index profile is more abrupt, and this opens up the possibilities of generating gradients by masks. The mask they use is, if I remember correctly, a silicon nitride or aluminum mask. So if you want to make an abrupt gradient profile this may be a way; but I think there are a few other ways to go. This type of research is important, so I wanted to inform you on this. Now the other material, arsenic trisulphide, which I indicated is a very important material because of the acousto-optic figure of merit, is one hundred times better than lithium niobate. If I also remember correctly there is some research going on in the University of Pennsylvania. I remember that they have observed this effect; after arsenic trisulphide is deposited on sapphire or silicon, if you expose it to light, at what wavelengths I don't remember, you can generate a large change in the index of refraction. So these are two cases I can think of. But as they say, they don't know what the mechanism is. So I can think of two very potentially useful materials in which you can induce large changes in the index of refraction. This type of research is important and also basic interactions are important. But I want to mention one thing. You people are familiar with the NSF-funded submicron fabrication center, now instituted at Cornell. NSF has supported the submicron fabrication center because of the fact that the cost for setting up equipment and facilities to make that kind of thing is so expensive, it is impossible for universities to do this. This will help people in device research to have the device fabricated there. So this brings up a question; in optical information processing research, is there a need for something similar to this; I don't mean that large a scale. I mean in a small scale that can be established by maybe ARO, or Air Force or Joint Services, to establish a facility that can fabricate some device so that each individual university does not have to repeat those things. You know that it is very difficult, as a modulator is very expensive, to get it from industry even if industry has it. But if the Joint Services can bring up some money to establish some useful facility, then I think that it will serve to expedite research among the university people. And I don't know whether there is such a need for optical information processing. [Chen Tsai]

JW: It is a very interesting question, certainly. While we are on the subject does anybody else want to comment on that? I think what he is saying is that some of us who are applying techniques but who aren't tied in with any manufacturer of devices, would love to get our hands on some things so that we can more closely relate to the realtime processing that the sponsoring people keep saying they want. You can have a technique in mind but if you don't have the devices it is pretty tough to try things out.

C: I've thought for some time now that it would be a great convenience for many of us doing optical processing work to have a facility available to us, perhaps by correspondence, to generate or produce exceptionally high quality computer generated optical elements, diffracting elements, maybe binary, maybe phase elements, maybe gray scale elements. There is a whole range of possibilities and there is a range of technologies that have not been exploited too much. [Bill Rhodes]

C: This Cornell facility might prove to be useful if they are really user oriented.

SS: I don't know much about the history of how the Cornell facility was developed. But if it does what it is supposed to do and if it works out, if the management can be worked out, I think it will be a very useful resource. I can see a lot of problems in managing it, like who is going to determine the priorities, who is going to pay for how it runs and so forth. But if something like that could be established, say for making optical transducers, that would be a very welcome resource.

JW: On the issue of computer generated masks, I think the most impressive facility that I have seen is Sing Lee's. His computer controlled scanners that he uses now for generating binary holograms are very impressive, and when you start talking about wanting to multiplex computer generated holograms, it is almost essential to have a facility like that. And they are rather expensive because you need a minicomputer, you need the scanner, etc.

SS: In fact, the integrated circuit people at Silicon Valley and other places have the finest mask making capability, but just try to get in and use them.

JW: Anybody want to respond?

C: I would like to make a comment about the Cornell submicron facility. The concept there is, by design, not really a job shop. It is a place where you can go and do research, principally a research place. As far

as specifying something and expecting them to do it, I don't think that will work. [Tom Gaylord]

SS: Let me say one other thing about this practical goal of how do we get a light valve in the hands of people who want to fool around with them. There is another way to look at this and I suppose we should look at the lessons of the Japanese. They have very strong university-industry-government cooperation and I guess we can spend hours and days talking about how they evolved to this state, why they are doing it and how they are very successful, at least from what we can see. But I think we need to move in this direction, and in order to do this it takes some very creative and argumentative salesmanship with the people in industry and in our specific case we have been working with Hughes-Malibu for awhile. We try to get light valves by the cheapest possible techniques, i.e. borrow them, but the trouble is we usually wind up with the rejects and they are constantly obsessed with questions of patent rights, who owns this and what if we invent something, who is going to get the rights and all such things like this. It takes a lot of creative, legal argumentation to try to convince these guys that we are really in the same boat and we are all trying to work toward a common goal. So if somehow this could be expedited by talking to people it would be very helpful.

C: There is currently a mechanism in the Army for an individual in the university to go to an Army lab and use their facilities. I tried for two years to get graduate students into my lab down in Huntsville, but I was unsuccessful, because no faculty member wants to release his graduate students from under his control for that length of time; I had three liquid crystal light valves, they are still down there, they run every day, they are available to anyone who wants to use them and I am sure those guys will welcome you to their laboratories; but you have to leave home. [Bob Guenther]

JW: That's part of the problem. How about loaning them to the labs, to the universities, that is where the people are?

C: No question that that was the major problem, to get the graduate students to come down to work for awhile. [Bob Guenther]

HW: Government laboratories can make equipment available to universities but the problem is that government laboratories have to purchase it from someplace and ultimately in Washington it has to be paid for. But for equipment which is going to be used or has been used to evaluate development systems, when the evaluation is completed, government ownership is maintained in the material. This material can then be transferred to universities, often on what amounts to indefinite loan, which is for all practical purposes equivalent to a gift. This is the way that very expensive pieces of equipment can be obtained, usually a year or so after they were available in industry. So you either get current lower quality equipment or, with the time delay, you can get the expensive equipment.

JW: One thing I hear Sandy saying is that there is a need for more cooperation, which the Japanese apparently are real good at, and sometimes we are not real good at.

C: I would like to address the general sensor question along that same line. If you look at what happens with industry in the United States, they feel that it is important to their industrial edge that they develop a group that is large enough so that they have a few people doing sensor work. Now, if the company is lucky enough, perhaps it has gifted enough people, to get a certain edge in sensors, I think that it is unrealistic to expect that a university group should get first cut on that kind of a device. It is fairly obvious, I think, that the main purpose of a meeting like this is that we meet some very outstanding people that are right at the leading edge of the device work right now. So we can go and talk to them and we can talk about how to get their devices, what their potentials are, and so on. There are a few university groups, and I'll advertise the Institute of Optics. One advantage of the Institute of Optics is that we are a large enough group that we can afford a person or so in one area and another person in another area, but even that I don't think is an answer. I don't think the government user facilities are an obvious end all, because one year, as we have seen through the past six or eight, one year it will be one device, and then another year there will be another one. Those things are changing, so if you try to grab the one today and say what it's going to be in five years, that is very touchy. The government labs build up a very elaborate facility, very, very, first class. I think it is naive to expect them to lend their liquid crystal which might be the heart of a 200,000 dollar installation. I have got a couple of nice installations in my lab if anyone wants the key. [Nick George]

C: It cost us a million and a half for a simple installation and a quarter of a million just to do some simple dielectric work on the PROM.

C: I was trying to level down in case the press was here. But if someone wants to come and work in my lab that is fine. If they want to borrow a piece on a long term thing and the students put their fingers on it, never mind, I am not interested. I think it is a good offer. [Nick George]

C: I would like to address the problem of getting spatial light modulators into the hands of the users, the systems people. In developing the microchannel spatial light modulator for example, we had lots of difficulties in trying to get a United States company to package the image intensifier part of it. We'd give them the crystal and microchannel plate and ask them just to package it so that we could employ a visible photo cathode, because as you know the bialkaline metals and trialkaline photocathodes decompose on exposure to air. So they have to be vacuum tight and they have to be sealed off. We wanted to test sensitivity to see if the design was really viable. We got quotations from companies for just the packaging like $20,000 to

$30,000 because we wanted only one. [Cardinal Warde]

R: That is very cheap.

C: Not if you have a $50,000 budget: So we got frustrated, we tried for almost a year to get United States companies to do it. These were companies like ITT, Varian, PMR. They all have the facilities to do that. Finally at one of the SPIE meetings I walked into one of the demonstration booths and ran across Hamamatsu Corporation; they had intensifier tubes and they were all excited. They said, "We don't understand how the device works, but it is the technology we are interested in," and I started calling their managers here and in Japan and I finally got them to agree upon a price. The first price was like the American companies, that is, $20,000. Then I said, "Listen, my budget has only 5,000 dollars for this kind of work," and they said, "Oh, we are still interested though, I'll tell you what we'll do, we'll do it on a best effort basis; we'll put some of our development money into it." And they did. They built the first device for us. They charged $5,000 and the thing worked, much to our dismay. We expected that they didn't know what we were doing, they didn't even test it. They built it and the president of their company put it in his pocket and shipped it all the way from Japan. They are now so interested in the device that they have decided to try and build it commercially and it may be that in the next year or two you might be able to buy one, but unfortunately it is going to say, "Made in Japan" instead of "Made in United States." The problem is that they don't really know how to build the device; if you want a device with high spatial resolution, high sensitivity, you probably have to work through me initially until they know what they are doing. They are just production engineers and they don't now have the ability to test the device. [Cardinal Warde]

C: This interaction between the universities and industries has been identified as a major problem in the USA and identified as something that DOD would like to see improved, but we don't have any solution, except that we would look favourably on this type of interaction. But I don't know how we generate this type of interaction. [Bob Guenther]

SS: At this point it has to be money.

R: No, it is not money, it is definitely not money. [Bob Guenther]

C: I would like to add one more comment. For those of you who are interested in acquiring a micro-channel spatial light modulator for use in the laboratories, where you can tolerate the fact that you might not have an optimum device, it turns out that the device is not really hard to fabricate in the laboratory. After all we can do it, and if you can live with low sensitivity, I shouldn't even say low sensitivity, because the sensitivity, even with excellent photo cathode is much better than other devices, you can get in contact with me and I can show you how to build it on your optical bench. You can store information for two weeks if you wanted to and you can cycle at 20Hz as we presently do. This is the sort of thing a graduate student with some assistance could do; so don't rule out the possibility, at least in my case, for interaction; because we are all in this together. [Cardinal Warde]

JW: The last line was "we are all in this together," and I think maybe we have to have a greater appreciation of that.

SS: I am in the university and if I wanted to talk to an industry about doing something cooperatively, at first they will say, "Go away, we know all the answers, we don't need you guys." But if I can go in and tell them about some applications for an existing device that they have not thought of, and sell them and show them that these things that they have sitting in their drawers, gathering dust, could be used for something new and useful, eventually you will find someone who will listen, and you just have to do an intensive sales job. And I think that is about the only way to do it. Just as Cardinal (Warde) said, you have to search around and find the right person and then get them one on one and convince them that it is to their advantage to ultimately listen to you and maybe he will let you borrow some of these devices out of his drawer.

JG: I should mention that NSF has had, I don't know for quite how long, a joint industry-university program in which at least up till recently they would provide funding to both organizations, usually through a subcontract through the university; but as of two or three weeks ago I was led to believe that the monetary contributions from NSF to the industry have now vanished and industry must now foot the bill for their part of the program and that may diminish the interest in the program.

C: Wouldn't they complain about patent rights and stuff like that? I work in Sandia laboratories which is managed by private industry but everything in it is owned by the government. But you know, we can offer real live money and you can't talk to these guys because of patent rights, etc. They're afraid someone is going to steal their super idea and make a million bucks on it. [Terry Stalker]

C: I think universities should have a different philosophy. I think universities should work on the basis of whatever we can get although we probably have a lower budget. But we should have to stretch the budget to do things better, probably not in direct competition with industry. There is no way we can compete with industry or government. It is probably fortunate that we don't have so much good equipment and we can probably do things which are different and which can get neat results. [H. K. Liu]

C: You have to be inventive, there is no question about that. There is such a big deal about the national good, how to get the most benefit for the nation in research or whatever, but I think you still have to argue

that basically if someone has a very good idea and is willing to implement it, they should have access to the equipment necessary. That does not mean you should be in competition with industry, but to hand a guy a light valve because he doesn't have 50,000 dollars, to make a pool of money available and get twenty light valves, and ten to circulate and get chewed up in the universities, or something like that, that is not really unreasonable. [Terry Stalker]

C: If you don't have a light valve, maybe you have to be forced to look at the major literature to see if the devices are available or materials are available to build your own. [H. K. Liu]

JW: I think this is a good point. In many cases it is the government labs that have the equipment and something that I have realized more lately is that they have a shortage of people because of some hiring freeze, etc; so obviously there is a need for more interaction between universities and government labs.

C: Even if a center was established, much like NSF has been establishing its regional facilities; you still have to leave your lab unless you were lucky enough to get it. And that is, in fact, what is wrong with the regional NSF user facilities, the guys who got them are in great shape. The guys who didn't get them hate them. [Bob Guenther]

HW: We can take students to work part time in the government laboratories, but you say that separates the students from the professor. We also have appointments which allow a professor to come to the government labs for three months during the normal summer break of the universities. Thus the student can be there part-time and the professor can be there for three months. For independent graduate students this may be a workable arrangement.

C: I hate to bring this one up. There is actually an even more formal arrangement called an IPA and you can go for two years. There are plenty of mechanisms for faculty members to interact with the government; graduate students to interact with the government; for industry members to interact with the government at the government labs and actually get paid. [Bob Guenther]

JW: Part of the problem here is that, in my experience, the way the universities work, the way the promotion and tenure system is geared, does not necessarily encourage people to go off and spend time in the laboratories when they should, in theory, be at home developing their own programs.

The workshop attendees discussed the need for this type of workshop and how potential future workshops could be set up. It was voted to hold similar workshops every two years, alternating with the Gordon Research Conference, beginning in 1983. Details of organizers, sponsors, and location were left open.

List of Participants

Dr. Carl C. Aleksoff
ERIM
P. O. Box 8618
Ann Arbor, Michigan 48103

Mr. Thomas J. Bicknell
Jet Propulsion Laboratory
Bldg. 183, Room #701
4800 Oak Grove Drive
Pasadena, California 91103

Prof. Steven K. Case
Dept. of Electrical Engineering
University of Minnesota
Minneapolis, Minnesota 55455

Prof. Stuart A. Collins, Jr.
Dept. of Electrical Engineering
Ohio State University
2015 Neil Avenue
Columbus, Ohio 43210

Prof. James M. Florence
Dept. of Electrical Engineering
University of Texas
Austin, Texas 78712

Prof. Thomas K. Gaylord
School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332

Prof. Nicholas George
Director, Institute of Optics
University of Rochester
Rochester, New York 14627

Dr. Mike Giles
White Sands Missile Range
Attn: STEWS-ID-T
White Sands Missile Range, NM 88002

Prof. Joseph W. Goodman
Dept. of Electrical Engineering
Stanford University
Stanford, California 94305

Dr. Bob D. Guenther
US Army Research Office
P. O. Box 12211
Attn: DRXRO-PH
Research Triangle Park, NC 27709

Prof. Marion O. Hagler
Dept. of Electrical Engineering
Texas Tech University
Lubbock, Texas 79409

Mr. Alan Huang
Information Systems Laboratory
Stanford Electronic Laboratories
Stanford University
Stanford, California 94305

Prof. Thomas F. Krile
Dept. of Electrical Engineering
Texas Tech University
Lubbock, Texas 79409

Dr. B. V. K. Vijaya Kumar
Dept. of Electrical Engineering
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

Dr. John N. Lee
Harry Diamond Laboratories
Branch 13200
2800 Powder Mill Road
Adelphi, Maryland 20783

Dr. Wai-Hon Lee
Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, California 94304

Dr. Robert D. Leighty
US Army Engineer Research Institute
Topographic Labs
Fort Belvoir, Virginia 22060

Prof. Emmett N. Leith
Electrical & Computer Engineering Dept.
University of Michigan
Ann Arbor, Michigan 48109

Prof. Hua-Kuang Liu
Dept. of Electrical Engineering
University of Alabama
University, Alabama 35486

Dr. Anthony Vander Lugt
Harris Corporation
P. O. Box 37
Melbourne, Florida 32901

Prof. Robert J. Marks II
Dept. of Electrical Engineering
University of Washington
Seattle, Washington 98195

Mr. Michael A. Monahan
Naval Ocean Systems Center
NOSC Code 811
San Diego, California 92152

Dr. John A. Neff
Program Manager
AFOSR/NE
Bolling AFB, Washington D.C. 20332

Prof. Thomas G. Newman
Dept. of Mathematics
Texas Tech University
Lubbock, Texas 79409

Prof. William M. Portnoy
Dept. of Electrical Engineering
Texas Tech University
Lubbock, Texas 79409

List of Participants Contd.

Prof. William T. Rhodes
School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332

Dr. William Sander
US Army Research Office
P. O. Box 12211
Attn: DRXRO-EL
Research Triangle Park, NC 27709

Prof. Alexander A. Sawchuk
Electrical Engineering Dept.
Image Processing Institute
University of Southern California
Los Angeles, California 90007

Mr. Rodolfo Segura
MS 470
Langley Research Center
Hampton, Virginia 23665

Dr. K. Terry Stalker
Sandia National Laboratories
Division 4426
P. O. Box 5800
Albuquerque, New Mexico 87185

Dr. William Stoner
Science Applications, Inc.
3 Preston Court
Bedford, Massachusetts 01730

Prof. Armand R. Tanguay, Jr.
Depts. of Electrical Engineering and
Materials Science
University of Southern California
Los Angeles, California 90007

Prof. Chen S. Tsai
Dept. of Electrical Engineering
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

Dr. Carl M. Verber
Battelle Columbus Laboratories
505 King Avenue
Columbus, Ohio 43201

Prof. John F. Walkup
Dept. of Electrical Engineering
Texas Tech University
Lubbock, Texas 79409

Prof. Cardinal Warde
Electrical Engineering & Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Mr. Harper Whitehouse
Naval Ocean Systems Center
NOSC Code 5303
San Diego, California 92152

END

DATE
FILMED